

This Page Is Inserted by IFW Operations
and is not a part of the Official Record

BEST AVAILABLE IMAGES

Defective images within this document are accurate representations of the original documents submitted by the applicant.

Defects in the images may include (but are not limited to):

- BLACK BORDERS
 - TEXT CUT OFF AT TOP, BOTTOM OR SIDES
 - FADED TEXT
 - ILLEGIBLE TEXT
 - SKEWED/SLANTED IMAGES
 - COLORED PHOTOS
 - BLACK OR VERY BLACK AND WHITE DARK PHOTOS
-
- GRAY SCALE DOCUMENTS

IMAGES ARE BEST AVAILABLE COPY.

**As rescanning documents *will not* correct images,
please do not report the images to the
Image Problem Mailbox.**

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 14, 2001, 13:48:40 ; Search time 1862.98 seconds
(without alignments)
1495.893 Million cell updates/sec

Title: US-09-292-437-2

Perfect score: 621
1 atgaaataatgacaaatcg.....tagctacagaagtaacataa 621

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 12496460 seqs, 2243815623 residues

Total number of hits satisfying chosen parameters: 24992920

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 10%

Listing first 45 summaries

Database : Pending_Patents_NA:*

1: /cgn2_6/ptodata/1/pna/US0877.COMB.seq:*
2: /cgn2_6/ptodata/1/pna/US0877.COMB.seq:*
3: /cgn2_6/ptodata/1/pna/US0877.COMB.seq:*
4: /cgn2_6/ptodata/1/pna/US0877.COMB.seq:*
5: /cgn2_6/ptodata/1/pna/US0877.COMB.seq:*
6: /cgn2_6/ptodata/1/pna/US0877.COMB.seq:*
7: /cgn2_6/ptodata/1/pna/US0877.COMB.seq:*
8: /cgn2_6/ptodata/1/pna/US0877.COMB.seq:*
9: /cgn2_6/ptodata/1/pna/US0877.COMB.seq:*
10: /cgn2_6/ptodata/1/pna/US0877.COMB.seq:*
11: /cgn2_6/ptodata/1/pna/US0877.COMB.seq:*
12: /cgn2_6/ptodata/1/pna/US0877.COMB.seq:*
13: /cgn2_6/ptodata/1/pna/US0877.COMB.seq:*
14: /cgn2_6/ptodata/1/pna/US0877.COMB.seq:*
15: /cgn2_6/ptodata/1/pna/US0877.COMB.seq:*
16: /cgn2_6/ptodata/1/pna/US0877.COMB.seq:*
17: /cgn2_6/ptodata/1/pna/US0877.COMB.seq:*
18: /cgn2_6/ptodata/1/pna/US0877.COMB.seq:*
19: /cgn2_6/ptodata/1/pna/US0877.COMB.seq:*
20: /cgn2_6/ptodata/1/pna/US0877.COMB.seq:*
21: /cgn2_6/ptodata/1/pna/US0877.COMB.seq:*
22: /cgn2_6/ptodata/1/pna/US0877.COMB.seq:*
23: /cgn2_6/ptodata/1/pna/US0877.COMB.seq:*
24: /cgn2_6/ptodata/1/pna/US0877.COMB.seq:*
25: /cgn2_6/ptodata/1/pna/US0877.COMB.seq:*
26: /cgn2_6/ptodata/1/pna/US0877.COMB.seq:*
27: /cgn2_6/ptodata/1/pna/US0877.COMB.seq:*
28: /cgn2_6/ptodata/1/pna/US0877.COMB.seq:*
29: /cgn2_6/ptodata/1/pna/US0877.COMB.seq:*
30: /cgn2_6/ptodata/1/pna/US0877.COMB.seq:*
31: /cgn2_6/ptodata/1/pna/US0877.COMB.seq:*
32: /cgn2_6/ptodata/1/pna/US0877.COMB.seq:*
33: /cgn2_6/ptodata/1/pna/US0877.COMB.seq:*
34: /cgn2_6/ptodata/1/pna/US0877.COMB.seq:*
35: /cgn2_6/ptodata/1/pna/US0877.COMB.seq:*
36: /cgn2_6/ptodata/1/pna/US0877.COMB.seq:*
37: /cgn2_6/ptodata/1/pna/US0877.COMB.seq:*
38: /cgn2_6/ptodata/1/pna/US0877.COMB.seq:*
39: /cgn2_6/ptodata/1/pna/US0877.COMB.seq:*
40: /cgn2_6/ptodata/1/pna/US0877.COMB.seq:*
41: /cgn2_6/ptodata/1/pna/US0877.COMB.seq:*
42: /cgn2_6/ptodata/1/pna/US0877.COMB.seq:*
43: /cgn2_6/ptodata/1/pna/US0877.COMB.seq:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Length	DB ID	Description
1	621	100.0	US-09-292-437-2	Sequence 2, Appli
2	621	100.0	US-08-781-986A-538	Sequence 538, App
3	621	100.0	US-08-956-171-538	Sequence 538, App
4	621	100.0	US-08-956-171B-538	Sequence 538, App
5	611.4	98.5	US-08-827-356-2678	Sequence 2678, Ap
6	611.4	98.5	US-09-611-529-2938	Sequence 2938, Ap
7	611.4	98.5	US-08-831-156A-108	Sequence 108, App
8	611.4	98.5	US-09-611-529-7449	Sequence 7449, App
9	572.2	92.1	US-06-038-697-735	Sequence 735, App
10	572.2	92.1	US-06-046-714-834	Sequence 714, App
11	534.2	86.0	US-09-620-608-849	Sequence 849, App
12	534.2	86.0	US-06-144-883-849	Sequence 849, App
13	340.2	54.8	US-09-450-969-1132	Sequence 1132, App
14	331.4	53.4	US-06-068-228-160	Sequence 160, App
15	331.4	53.4	US-06-038-081-348	Sequence 348, App
16	331.4	53.4	US-06-046-953-355	Sequence 355, App
17	194	31.2	US-06-206-848-39	Sequence 39, Appl
18	194	31.2	US-06-242-578-165	Sequence 165, App
19	194	31.2	US-06-257-931-163	Sequence 163, App
20	194	31.2	US-06-253-625-165	Sequence 165, App
21	194	31.2	US-06-269-308-165	Sequence 165, App
22	75.2	12.1	US-09-404-520-13349	Sequence 13349, A
23	69.2	11.1	US-06-211-750-24644	Sequence 24644, A
24	68.8	11.1	US-09-397-761A-2720	Sequence 2720, Ap
25	68.8	11.1	US-09-641-529-376	Sequence 376, App
26	68.4	11.0	US-06-207-458-143594	Sequence 143594,
27	68.2	11.0	US-09-397-761A-2719	Sequence 2719, Ap
28	68.2	11.0	US-09-641-529-411	Sequence 411, App
29	66.6	10.7	US-09-397-761A-2717	Sequence 2717, Ap
30	66.6	10.7	US-09-641-529-380	Sequence 380, App
31	66.2	10.7	US-06-211-750-71125	Sequence 71125, A
32	66	10.6	US-09-565-306-29077	Sequence 29077, A
33	66	10.6	US-06-207-458-58087	Sequence 58087, A
34	66	10.6	US-09-397-761A-2710	Sequence 2710, Ap
35	66	10.6	US-09-641-529-448	Sequence 448, App
36	65.6	10.6	US-06-207-458-88810	Sequence 88810, A
37	65.6	10.6	US-09-404-520-13067	Sequence 13067, A
38	65.4	10.5	US-08-466-194-14	Sequence 14, Appl
39	65	10.5	US-09-397-761A-2715	Sequence 2715, Ap
40	65	10.5	US-09-641-529-680	Sequence 680, App
41	65	10.5	US-06-212-659-6	Sequence 6, Appli
42	65	10.5	US-06-208-020-72	Sequence 72, Appli
43	65	10.5	US-06-209-043-95	Sequence 95, Appli
44	65	10.5	US-06-213-172-35	Sequence 35, Appli
45	65	10.5	US-06-213-173-35	Sequence 35, Appli

ALIGNMENTS

RESULT 1
US-09-292-437-2
Sequence 2, Application US/09292437
GENERAL INFORMATION:
APPLICANT: Olaf Schneewind
APPLICANT: Sarkis Mazmanian
APPLICANT: Gwen Liu
APPLICANT: Hung Ton-That
TITLE OF INVENTION: IDENTIFICATION OF SORTASE GENE
FILE REFERENCE: 510015.213
CURRENT APPLICATION NUMBER: US/09/292.437
CURRENT FILING DATE: 1999-04-15
NUMBER OF SEQ ID NOS: 36
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 2
LENGTH: 621
TYPE: DNA
ORGANISM: Staphylococcus aureus
US-09-292-437-2

Query Match 100.0%; Score 621; DB 55; Length 621;
Best Local Similarity 100.0%; Pred. No. 9.8e-107;
Matches 621; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 atgaaaaatggaacacatcgattatgacacatcgctgctgctgctatctatctagtgca 60
1 atgaaaaatggaacacatcgattatgacacatcgctgctgctgctatctatctagtgca 60
Db 1 atgaaaaatggaacacatcgattatgacacatcgctgctgctgctatctatctagtgca 60
QY 61 gcatattgttctgtaaacacacatcgatgataatcttccagtaagaataagatgaa 120
61 gcatattgttctgtaaacacacatcgatgataatcttccagtaagaataagatgaa 120
Db 61 gcatattgttctgtaaacacacatcgatgataatcttccagtaagaataagatgaa 120
QY 121 aagattgaacacatgataaataatgtaaaagaaacagcgagtaagaataaagacgaa 180
121 aagattgaacacatgataaataatgtaaaagaaacagcgagtaagaataaagacgaa 180
Db 121 aagattgaacacatgataaataatgtaaaagaaacagcgagtaagaataaagacgaa 180
QY 181 gctaaacctcaattccgaaagataatcgaaagtcgacgctatctgtaaatccagat 240
181 gctaaacctcaattccgaaagataatcgaaagtcgacgctatctgtaaatccagat 240
Db 181 gctaaacctcaattccgaaagataatcgaaagtcgacgctatctgtaaatccagat 240
QY 241 gctgataatgaagacagatataccagacagacacacacacacacacacacacac 300
241 gctgataatgaagacagatataccagacagacacacacacacacacacacacac 300
Db 241 gctgataatgaagacagatataccagacagacacacacacacacacacacacac 300
QY 301 gtaagcttcgagaagaataatgtaacatcgatgataatcttcaatttcgagagcac 360
301 gtaagcttcgagaagaataatgtaacatcgatgataatcttcaatttcgagagcac 360
Db 301 gtaagcttcgagaagaataatgtaacatcgatgataatcttcaatttcgagagcac 360
QY 361 acttcattgacgctccgaactatcaatttcaaatcttaagcagcacaataagatgagt 420
361 acttcattgacgctccgaactatcaatttcaaatcttaagcagcacaataagatgagt 420
Db 361 acttcattgacgctccgaactatcaatttcaaatcttaagcagcacaataagatgagt 420
QY 421 atgggtgactttaaagtctgtaagaacacgtaagataataatgtaagaatgagat 480
421 atgggtgactttaaagtctgtaagaacacgtaagataataatgtaagaatgagat 480
Db 421 atgggtgactttaaagtctgtaagaacacgtaagataataatgtaagaatgagat 480
QY 481 gtaagactcagatgtagagatctagatgaaacaaaagtaagaataaacaattaca 540
481 gtaagactcagatgtagagatctagatgaaacaaaagtaagaataaacaattaca 540
Db 481 gtaagactcagatgtagagatctagatgaaacaaaagtaagaataaacaattaca 540
QY 541 ttaattacttgatgataatcaatgaaagacagcgcttggtggaataaagcttaaatctt 600
541 ttaattacttgatgataatcaatgaaagacagcgcttggtggaataaagcttaaatctt 600
Db 541 ttaattacttgatgataatcaatgaaagacagcgcttggtggaataaagcttaaatctt 600
QY 601 gtagctacagaagatcaataa 621
601 gtagctacagaagatcaataa 621
Db 601 gtagctacagaagatcaataa 621

RESULT 2
US-08-781-986A-538/c
Sequence 538, Application US/08781986A

GENERAL INFORMATION:
APPLICANT: Charles Kunsch
TITLE OF INVENTION: Staphylococcus aureus Polynucleotides and Sequences
NUMBER OF SEQUENCES: 5255
CORRESPONDENCE ADDRESS:
ADDRESSEE: Human Genome Sciences, Inc.
STREET: 9410 Key West Avenue
CITY: Rockville
STATE: Maryland
COUNTRY: USA
ZIP: 20850
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 3.50 inch, 1.4mb storage
COMPUTER: HP Vectra 486/33
OPERATING SYSTEM: MSDOS version 6.2
SOFTWARE: ASCII Text
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/781, 986A
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Benson, Bob
REGISTRATION NUMBER: 30,446
REFERENCE/DOCKET NUMBER: PB248PP
TELECOMMUNICATION INFORMATION:
TELEPHONE: (301) 309-8504
TELEFAX: (301) 309-8512
INFORMATION FOR SEQ ID NO: 538:
SEQUENCE CHARACTERISTICS:
LENGTH: 3733 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
US-08-781-986A-538

Query Match 100.0%; Score 621; DB 11; Length 3733;
Best Local Similarity 100.0%; Pred. No. 1.2e-106;
Matches 621; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 atgaaaaatggaacacatcgattatgacacatcgctgctgctgctatctatctagtgca 60
730 ATGAAAAATGGAACAAATCGATTATGCAATCGCTGCTGCTGCTATCTTATCTTAGTGA 671
Db 61 gcatattgttctgtaaacacacatcgatgataatcttccagtaagaataagatgaa 120
670 GCATATTGTTGTTGCTAAACCAATATCGATGATATCTTACGATTAAGATAAGATGAA 611
QY 121 aagattgaacacatgataaataatgtaaaagaaacagcgagtaagaataaagacgaa 180
610 AAGATTGAACATATGATTAATAATGTAAAGACAGCGAGTAAAGATAAGATAAGACGAA 551
QY 181 gctaaacctcaattccgaaagataatcgaaagtcgacgctatctgtaaatccagat 240
550 GCTAAACCTCAAAATTCGAAAGATAATCGAATGCGAGGCTATATTGAATTCAGAT 491
Db 490 GCTGATATTAAAGAACAGTATATCCAGACACACACACCTGAACATTAATAGAGCT 431
QY 241 gctgataatgaagacagatataccagacagacacacacacacacacacacacac 300
490 GCTGATATTAAAGAACAGTATATCCAGACACACACACCTGAACATTAATAGAGCT 431
Db 241 gctgataatgaagacagatataccagacagacacacacacacacacacacacac 300
QY 301 gtaagcttcgagaagaataatgtaacatcgatgataatcttcaatttcgagagcac 360
430 GTAAGCTTTGAGAGAAAGAAATGAAATCACTAGATGATCAAAATTAATTTGACAGGAC 371
Db 361 acttcattgacgctccgaactatcaatttcaaatcttaagcagcacaataagatgagt 420
QY 370 ACTTTCATTGACGCTCCGAACATCAATTTACAAATCTTAAGACAGCCAAAAGATGAT 311
421 atgggtgactttaaagtctgtaagaacacgtaagataataatgtaagaatgagat 480

Db 310 ATGGTGTCTTAAAGTTGGTAATGAAACACGTAAGTATAAATACAAAGTAAAGAGAT 251
Qy 481 gtaagcctacagatgtagagttctagatgacacaaagtaaaagataaacaatlaaca 540
Db 250 GTTAAGCCTACAGATGTAGAGTTCTAGATGACAAAAAGTAAGATTAACATTTACA 191
Qy 541 ttaattactgtgatgattacaaatgaaagacagcgcttgggaaacgttaaatcttt 600
Db 190 TTAATTACTGTGATGATTACAAATGAAAAAGACAGCGTTGGGAAAAAGTAAATCTTT 131
Qy 601 gtaagctacagaagtaacaataa 621
Db 130 GTAGCTACAGAAGTCAATAATA 110

RESULT 3

US-08-956-171-538/c
Sequence 538, Application US/08956171
GENERAL INFORMATION:
APPLICANT: Charles Kunsch
TITLE OF INVENTION: Staphylococcus aureus Polynucleotides and Sequences
NUMBER OF SEQUENCES: 5255
CORRESPONDENCE ADDRESS:
ADDRESSEE: Human Genome Sciences, Inc.
STREET: 9410 Key West Avenue
CITY: Rockville
STATE: Maryland
COUNTRY: USA
ZIP: 20850
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage
COMPUTER: HP Vectra 486/33
OPERATING SYSTEM: MSDOS version 6.2
SOFTWARE: ASCII Text
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/956.171
FILING DATE:
CLASSIFICATION: 424
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Benson, Bob
REGISTRATION NUMBER: 30,446
REFERENCE/DOCKET NUMBER: PB248BP
TELECOMMUNICATION INFORMATION:
TELEPHONE: (301) 309-8504
TELEFAX: (301) 309-8512
INFORMATION FOR SEQ ID NO: 538:
SEQUENCE CHARACTERISTICS:
LENGTH: 3733 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
US-08-956-171-538

Query Match

Best Local Similarity 100.0%; Score 621; DB 13; Length 3733;
Best Local Similarity 100.0%; Pred. No. 1.2e-106;
Matches 621; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 atgaaaaatgacaaatcgatataatgacaaatcgatgtagtactactactactgagca 60
Db 730 ATGAAAAATGACAAATCGATATATGACAAATCGCTGGTGTGCTACTTATCTTACTG 671
Qy 61 gcatattgttctgtaaacacacatacgataatattcttcacgataaagataaagatgaa 120
Db 670 GCATATTGTGCTGTAACCAATCATATATATCTTACACGATAAAGATTAAGATGAA 611
Qy 121 aagattgacaaatgataaataatgtaaaagacagcgagagtaaaagataaagaagcaaa 180
Db 610 AAGATTGAACATATGATTAATAATGTAAAGAACAGCGCAGTAAGATTAATAAGACAGCA 551

Qy 181 gctaaacctcaaatctccgaaagataaaltcgaaagtggcaggtatattgaaatccagat 240
Db 550 GCTAAACCTCAAAATTCGGAAGATTAATCGAAAGTGGCAGGCTATATTGAATTCAGAT 491
Qy 241 gctgatattaaagaaccagatataatccagagccagcaaccctgaaacatlaaatgaggt 300
Db 490 GCTGATATTAAAGAACCACTATATTCGAGACGACGAAACCTGAAACATTAATAGAGCT 431
Qy 301 gtaagcttgcagaagaaatgaatcacatagatgataaataatltcaaltgcaagacac 360
Db 430 GTAGCTTGACAGAAATATGATCAGATGATCAAAATTTTCATTTCAGAGACAC 371
Qy 361 acttcaattgacgcgtccgaactcaatlaacaatcttaagcagccaaaaaggtagt 420
Db 370 ACTTCATGTGACCGCCGCAACTATCAATTTACAAATCTTAAGACACCAAAAAAGTAGT 311
Qy 421 atgggtactttaaagttggtatgaaacacgttaagtataaataatgcaagtataagat 480
Db 310 ATGGTGTCTTAAAGTTGGTAATGAAACACGTAAGTATAAATGACAAAGTAAAGAGAT 251
Qy 481 gtaagcctacagatgtagagttctagatgacacaaagtaaaagataaacaatlaaca 540
Db 250 GTTAAGCCTACAGATGTAGAGTTCTAGATGACAAAAAGTAAGATTAACATTTACA 191
Qy 541 ttaattactgtgatgattacaaatgaaagacagcgcttgggaaacgttaaatcttt 600
Db 190 TTAATTACTGTGATGATTACAAATGAAAAAGACAGCGCTTGGGAAAAAGTAAATCTTT 131
Qy 601 gtaagctacagaagtaacaataa 621
Db 130 GTAGCTACAGAAGTCAATAATA 110

RESULT 4

US-08-956-171B-538/c
Sequence 538, Application US/08956171B
GENERAL INFORMATION:
APPLICANT: Charles Kunsch
Gill H. Choi
Patrick S. Dillon
Craig A. Rosen
Steven C. Barash
Michael R. Fannon
TITLE OF INVENTION: Staphylococcus aureus Polynucleotides and Sequences
NUMBER OF SEQUENCES: 5255
CORRESPONDENCE ADDRESS:
ADDRESSEE: Human Genome Sciences, Inc.
STREET: 9410 Key West Avenue
CITY: Rockville
STATE: Maryland
COUNTRY: USA
ZIP: 20850
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage
COMPUTER: HP Vectra 486/33
OPERATING SYSTEM: MSDOS version 6.2
SOFTWARE: ASCII Text
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/956.171B
FILING DATE: 20-Oct-1997
CLASSIFICATION: <Unknown>
PRIOR APPLICATION NUMBER:
APPLICATION NUMBER: 60/009 861
FILING DATE: January 5, 1996
APPLICATION NUMBER: 08/781,986
FILING DATE: January 3, 1997
ATTORNEY/AGENT INFORMATION:
NAME: Hoover, Kenley K.
REGISTRATION NUMBER: 40,302
REFERENCE/DOCKET NUMBER: PB248P1
TELECOMMUNICATION INFORMATION:
TELEPHONE: (301) 610-5790
TELEFAX: (301) 309-8439

```

? INFORMATION FOR SEQ ID NO: 538:
?     SEQUENCE CHARACTERISTICS:
?         LENGTH: 3733 base pairs
?         TYPE: nucleic acid
?         STRANDEDNESS: double
?         TOPOLOGY: linear
?     SEQUENCE DESCRIPTION: SEQ ID NO: 538:
US-08-956-171B-538

```

Query Match	100.0%	Score 621	DB 13	Length 3733
Best Local Similarity	100.0%	Pred. No. 1	2e-106	
Matches 621; Conservative	0	Mismatches	0	Gaps 0

QY	1	atgaaaataatgcaaaaatcgaattatatacaaacgcctggctggtgacttaaccatgtgca	60
Db	730	ATGAAAATAATGCAAAATCGATTATATGACAAATCGCTGTGTGGTACTTATCTTGTGCGCA	671
QY	61	gcataattgcttctgtaaacacacatcatcgataaattactcttcacgtataaagataagtgaa	120
Db	670	GCATATTGTGTGCTTAAACCACATATCGATATATATATCTTCACGATAAGATTAAGATGAA	611
QY	121	aagattigaacacatctgataaaaatgataaagaacagccggttaaagataaagtcagcaa	180
Db	610	AAGATTGCAACATATGTATATAAAATGCTAAAGAACGCGCATGAAGCTATTAAGAACACACAA	551
QY	181	gctaaacctcaaatctcgaaaagataaatacgaagctggcaggtctatctgaattccagat	240
Db	550	GCTAAACCTCAAAATTCGGAAGATTAATTCGAANATGCGACGGCTATATTGAAATTCAGAT	450
QY	241	gcctgatattaaagaacagatataccagagaccgaacacctgaacaattaaataagat	300
Db	490	GCTATATTAAAGAACAGATATATCCAGGACACACACTGTACATTAATTAATAGAGT	431
QY	301	gtaagccttgcagaagaagaatgaaatgcactgataatacaaaatatttcaattctgacgacac	360
Db	430	GTAAGCTTTGCGAGAGAAATGAAATGAAATCTCTACATGATATAAAATATTTCATTTGCGGGCAC	371
QY	361	aacttcatltgacccgtccgaactacaaatttacaatalttlaagcagccaaaagaagctagt	420
Db	370	ACTTTCATGTGACCGTCGGAACCTTCAATTTNCAAAATCTTAAAGACGCCAAAAAAGTAGT	311
QY	421	atgctgcttaacttaaaagtttgttaataagaaacacgctgaagtataataatgacagctataagat	480
Db	310	ATGCTGTACTTTAAAGTTGTGTAATGAAGAACGTTAAGTATTAATAATGACAAAGTATAAGAGAT	251
QY	481	gttaagcctacagaatgttagagagttctatgataagaaacaaaagttaagataaacatataca	540
Db	250	GTTTAAGCCTACAGATGTATGAGAGTTCTTGATGACAAAAAGGTAAAGATTAACATTTTACA	191
QY	541	tttaattactgtgatagtatacaatlgaaaagacagcgtctgggaaaaagctaaatcttc	600
Db	190	TTAATTACTGTGTGATGATTACAATGAAAGACAGCGCTTTGGGAAAAACGTAAATCTTT	131
QY	601	gttaactacagaagatcaataaa 621	
Db	130	GTACTACAGAAGTCAATAAT 110	

RESULT 5
 US-08-827-356-2678
 ? Sequence 2678, Application US/08827356
 ? GENERAL INFORMATION:
 ? APPLICANT: George H. Shimer, Jr.
 ? APPLICANT: George H. Miller
 ? APPLICANT: Roberta S. Hare
 ? APPLICANT: Karen J. Shaw
 ? TITLE OF INVENTION: STAPHYLOCOCCUS AUREUS RELATED
 ? TITLE OF INVENTION: COMPOSITIONS AND METHODS
 ? NUMBER OF SEQUENCES: 574
 ? CORRESPONDENCE ADDRESS:
 ? ADDRESSEE: Schering-Plough Corporation
 ? STREET: 2000 Galloping Hill Road

```

1      CITY: Kenilworth
2      STATE: New Jersey
3      COUNTRY: USA
4      ZIP: 07033-0530
5
6      COMPUTER READABLE FORM:
7      MEDIUM TYPE: Floppy disk
8      COMPUTER: IBM, PC compatible
9      OPERATING SYSTEM: PC-DOS/MS-DOS
10     SOFTWARE: Patentin
11
12     CURRENT APPLICATION DATA:
13     FILING APPLICATION NUMBER: US/08/827,356
14     FILING DATE: 01-APR-1997
15     PRIOR APPLICATION DATA:
16     APPLICATION NUMBER: 60/014,477
17     FILING DATE: 01-APR-1996
18     APPLICATION NUMBER: 60/016,743
19     FILING DATE: 02-MAY-1996
20     APPLICATION NUMBER: 60/020,016
21     FILING DATE: 14-JUN-1996
22
23     INFORMATION FOR SEQ. ID NO: 2678:
24
25     SEQUENCE CHARACTERISTICS:
26
27     LENGTH: 624 base pairs
28     TYPE: nucleic acid
29     STRANDEDNESS: double
30     TOPOLOGY: circular
31
32     MOLECULE TYPE: DNA (genomic)
33
34     HYPOTHETICAL: NO
35
36     ANTI-SENSE: NO
37
38     ORIGINAL SOURCE:
39
40     ORGANISM: Staphylococcus aureus
41
42     FEATURE:
43
44     NAME/KEY: misc_feature
45     LOCATION: 1...624
46
47     US-08-827-356-2678

```

Query Match	98.5%	Score 611.4	DB 12	Length 624
Best Local Similarity	99.0%	Pred. NO. 6.1e-105		
Matches 615; Conservative	0	Mismatches 6	Indels 0	Gaps 0

OY	1	atgaaanaatgacaaatcgaattcaatcaacaatgcgtgtgtgtgtaacttatccatgtagca	60
Db	4	ATGAAAANAATGACAAATCGATTATGACAAATCCCTGCTGTAGTACTTATCTTAGTGGCA	63
OY	61	gcataattgtttctgtaaacacacatcatcgataaatacttccagataaagaataaagttaa	120
Db	64	GCATATTGTGTGGTAAACCACATATGATGATATATATCTTCACGATAAGATTAAGATGAA	123
OY	121	aagattggaacaattgtataaaatgttaagaacacggcgttaagaataaagaacgacaa	180
Db	124	AGATTGTGACAAATATGATATAAATATGTAAMAAAGCGGTAAAGCAATTAAGCACAA	183
OY	181	gtlaaaccttcaaatctcgaaagaataaactcgaaagttgcagcttatgtgaatctcagat	240
Db	184	GCTAAACCTCAAAATCCGAAGATTAATCAAAAGTGGCAGGCTATTTGTGAATTCAGAT	243
OY	241	gctgatataaagaacgaatataccgagaccgaacacctgtgaacaaataaataagaggt	300
Db	244	GCTGATTTATTAAGAACCGTATATCCAGGACCCACCACTCGTAACAAATTAATAAGGgt	303
OY	301	gtaagctttgcagaagaanaatgaaatcaactagatgatacaaaatatttcaatttgcagagac	360
Db	304	GTAAGCTTTGCGAAGAAAGAAATGATTCACATGATATCAAAATTTTCAATTGGCAGGAC	363
OY	361	actttcaattgacggttcggaactatcaatttacaatctttaaagcagccaaaaaagttagt	420
Db	364	ACTTTCAATTGACCGCTCCGCAACTATCAATTTTAACAATCTTTAAAGCAGCGCAAAAAAGGTAGT	423
OY	421	atggtgtactttaaagttgtgtaatgaaacacgtaagataaanaatgaaagttataaagat	480
Db	424	ATGGGTACTTTAAAGTTGGTAAATGAAACACGTAAAGTATTAATAATGACAAAGTATTAAGGAT	483
OY	481	gttaagcctcaagaatgttagaggtctatagatgaaacaaaagttaaagataaacaatttaca	540

```
Db 484 GTTAAGCCACAGATGTAGAACTTCTAGATGAACAAAAAGTAAAGATAAACCAATTACA 543
OY 541 ttaattacttctgtagatgatacaataaagacagcgctttgggaaacgtaaatcttt 600
Db 544 TTAATTACTTGTGATGATTACATGAAAAAGACAGCGCTTTGGGAAAAACGTAAAACTCTT 603
OY 601 gtagctacagaagtcataataa 621
Db 604 GTAGCTACAGAACTCAATAATA 624

RESULT 6
US-09-611-529-2938
: Sequence 2938, Application US/09611529
: GENERAL INFORMATION:
: APPLICANT: George H. Shimer, Jr.
: APPLICANT: George H. Miller
: APPLICANT: Robert S. Hare
: APPLICANT: Karen J. Shaw
: TITLE OF INVENTION: Staphylococcus aureus Related Compositions and Methods
: FILE REFERENCE: 1034/1C963JUS1
: CURRENT APPLICATION NUMBER: US/09/611,529
: PRIOR FILING DATE: 2000-06-30
: PRIOR APPLICATION NUMBER: US 09/417,811
: PRIOR FILING DATE: 1999-10-14
: PRIOR APPLICATION NUMBER: US 09/353,718
: PRIOR FILING DATE: 1999-07-14
: PRIOR APPLICATION NUMBER: US 09/266,557
: PRIOR FILING DATE: 1999-03-11
: PRIOR APPLICATION NUMBER: US 09/266,556
: PRIOR FILING DATE: 1999-03-11
: PRIOR APPLICATION NUMBER: US 09/266,555
: PRIOR FILING DATE: 1999-03-11
: PRIOR APPLICATION NUMBER: US 09/266,542
: PRIOR FILING DATE: 1999-03-11
: PRIOR APPLICATION NUMBER: US 09/266,541
: PRIOR FILING DATE: 1999-03-11
: PRIOR APPLICATION NUMBER: US 09/037,934
: PRIOR FILING DATE: 1998-03-10
: PRIOR APPLICATION NUMBER: US 09/036,720
: PRIOR FILING DATE: 1998-03-06
: PRIOR APPLICATION NUMBER: US 09/036,338
: PRIOR FILING DATE: 1998-03-06
: PRIOR APPLICATION NUMBER: US 09/036,334
: PRIOR FILING DATE: 1998-03-06
: PRIOR APPLICATION NUMBER: US 09/036,221
: PRIOR FILING DATE: 1998-03-06
: PRIOR APPLICATION NUMBER: US 09/036,137
: PRIOR FILING DATE: 1998-03-06
: PRIOR APPLICATION NUMBER: US 09/036,082
: PRIOR FILING DATE: 1998-03-06
: PRIOR APPLICATION NUMBER: US 09/036,081
: PRIOR FILING DATE: 1998-03-06
: PRIOR APPLICATION NUMBER: US 09/036,079
: PRIOR FILING DATE: 1998-03-06
: PRIOR APPLICATION NUMBER: US 09/035,913
: PRIOR FILING DATE: 1998-03-06
: PRIOR APPLICATION NUMBER: US 09/035,744
: PRIOR FILING DATE: 1998-03-06
: PRIOR APPLICATION NUMBER: US 08/827,356
: PRIOR FILING DATE: 1997-04-01
: PRIOR APPLICATION NUMBER: US 08/831,156
: PRIOR FILING DATE: 1997-04-01
: PRIOR APPLICATION NUMBER: US 60/014,477
: PRIOR FILING DATE: 1996-04-01
: PRIOR APPLICATION NUMBER: US 60/016,743
: PRIOR FILING DATE: 1996-05-02
: PRIOR APPLICATION NUMBER: US 60/020,016
: PRIOR FILING DATE: 1996-06-14
: NUMBER OF SEQ ID NOS: 7451
: SEQ ID NO 2938
: LENGTH: 624
```

```
: TYPE: DNA
: ORGANISM: Staphylococcus aureus
: US-09-611-529-2938

Query Match      98.5%; Score 611.4; DB 22; Length 624;
Best Local Similarity 99.0%; Pred. No. 6,1e-105;
Matches 615; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

OY 1 atgaaataatgacaaatctgataatgaacaatcgcgtggtgtgtacttctctatgtgca 60
Db 4 atgaaataatgacaaatctgataatgaacaatcgcgtggtgtgtacttctctatgtgca 63
OY 61 gcatattgttgcacaaacacatcgtataattatcttcacgataaagaatgaaga 120
Db 64 gcatattgttgcacaaacacatcgtataattatcttcacgataaagaatgaaga 123
OY 121 aagattgaacaaatgataaataatgaaagacagcgagtaagaatgaagaacagca 180
Db 124 aagattgaacaaatgataaataatgaaagacagcgagtaagaatgaagaacagca 183
OY 181 gctaaacctcaaatccgaaagataaattcgaagtgccaggtatattgaattccagat 240
Db 184 gctaaacctcaaatccgaaagataaattcgaagtgccaggtatattgaattccagat 243
OY 241 gctgataattaaagaccagatatatccagcagcagcaaccctgaacaattaaatgaagt 300
Db 244 gctgataattaaagaccagatatatccagcagcagcaaccctgaacaattaaatgaagt 303
OY 301 gtaagcttgcagaagaaatgaatcactagatgatacaataattcaattgcagagcac 360
Db 304 gtaagcttgcagaagaaatgaatcactagatgatacaataattcaattgcagagcac 363
OY 361 acttcattgaccggtccgaaactatcaattacaattcgaagcagcagcagcagcagcagc 420
Db 364 acttcattgaccggtccgaaactatcaattacaattcgaagcagcagcagcagcagcagc 423
OY 421 atggtactttaaagtttgtaatgaacacgtaagtataaattgacagtaagaagt 480
Db 424 atggtactttaaagtttgtaatgaacacgtaagtataaattgacagtaagaagt 483
OY 481 gttaaagctacagatgtagagttctagaatgaacaaagagtaagaatgaataaataa 540
Db 484 gttaaagctacagatgtagagttctagaatgaacaaagagtaagaatgaataaataa 543
OY 541 ttaattacttctgtagatgatacaataaagacagcgctttgggaaacgtaaatcttt 600
Db 544 ttaattacttctgtagatgatacaataaagacagcgctttgggaaacgtaaatcttt 603
OY 601 gtagctacagaagtcataataa 621
Db 604 gtagctacagaagtcataataa 624

RESULT 7
US-08-831-156A-108
: Sequence 108, Application US/08831156A
: GENERAL INFORMATION:
: APPLICANT: George H. Shimer, Jr.
: APPLICANT: George H. Miller
: APPLICANT: Robert S. Hare
: APPLICANT: Karen J. Shaw
: TITLE OF INVENTION: NUCLEOTIDE SEQUENCE OF A
: TITLE OF INVENTION: STAPHYLOCOCCUS AUREUS
: TITLE OF INVENTION: GENOME AND RELATED
: TITLE OF INVENTION: COMPOSITIONS AND METHODS
: NUMBER OF SEQUENCES: 112
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Schering-Plough Corporation
: STREET: 2000 Galloping Hill Road
: CITY: Kenilworth
: STATE: New Jersey
: COUNTRY: USA
```

```
ZIP: 07033-0530
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/831,156A
FILING DATE: 01-APR-1997
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 60/014,477
FILING DATE: 01-APR-1996
APPLICATION NUMBER: 60/016,743
FILING DATE: 02-MAY-1996
APPLICATION NUMBER: 60/020,016
FILING DATE: 14-JUN-1996
INFORMATION FOR SEQ ID NO: 108:
SEQUENCE CHARACTERISTICS:
LENGTH: 91077 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: circular
MOLECULE TYPE: DNA (genomic)
HYPOTHETICAL: NO
ANTI-SENSE: NO
ORIGINAL SOURCE:
ORGANISM: Staphylococcus aureus
FEATURE:
NAME/KEY: misc_feature
LOCATION: 1...91077
US-08-831-156A-108
```

```
Query Match          98.5%  Score 611.4:  DB 12:  Length 91077:
Best Local Similarity 99.0%:  Pred. No 1.2e-104:
Matches 615; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

QY      1  atgaaaaatgagacaaatcgatgaatgacaatcgctgtgtgtgactatcctagtggca  60
Db 15161 ATGMAAAATGAGCAAAATCGATTATGATGACATCGCTGTGTGACTTATCTGATGGCA 15220

QY      61  gcatattgttctgaaccacacatacgcataatcttccagataaagataagatgaa  120
Db 15221 GCATATTGTGTTGCTAAACCAATATGATGATATCTTCCGATTAAGATAAGATGAA 15280

QY     121  aagatgaacaataatgataaataatgataaagacagcgatgaagataaagacagcaa  180
Db 15281 AAGATTGAACAATATGATGATTAATAATGTAAGAACAGCGGATTAAGACAAATAGCAGCAA 15340

QY     181  gctaaacctcaattccgaaagataatcgaaagtgacaggtatattgaaattccagat  240
Db 15341 GCTAAACCTCAAAATTCGAAAGATAATCAAAAGTGCGACGCTATTGTAATTCAGAT 15400

QY     241  gctgatatgaagaacagcatatccagagccagcaacacctgtaacatttaagatgag  300
Db 15401 GCTGATATTAAAGACAGATATATCCAGACCCAGCAACACCTGAACATTAATTAAGAGGT 15460

QY     301  gtaagcttgcagaagaataatgaatcaactagatgatacaaaatattcgaatgcagacac  360
Db 15461 GTAAGCTTGCAGAGAAATGATCACTAGATGATCAAAATATTCTCAATTGCGAGCAC 15520

QY     361  acctcatgacgctccgaactcaattacaattctaaagcagcgaacaaagatgag  420
Db 15521 ACTTCAATGACGCTCGAATCATCAATTTAACAATCTTAAAGCAGCAAAAAGGTAGT 15580

QY     421  atggtgtactttaagtgtaataatgaacacgtaagataataaataagcaagataagat  480
Db 15581 ATGGTGTACTTTAAAGTTGCTAATGAACACGTAAGTATTAATAAGCAAGTATTAAGAT 15640

QY     481  gtttaagcctaagaatgtagagatctagatgaacaaaaagttaagataaacaattaca  540
Db 15641 GTTAAAGCCAACAGTATGAAGATTCTAGATGAACAAAAAGTAAAGTAAACATTAACA 15700
```

```
QY      541  ttaattactgtatgattacaatgaaagacagcgcttgggaaaaacgtaaaatcctt  600
Db 15701 TTAATTACTGTGATGATTACAATGAAGAACAGCGCTTTGGAAAAACGTAAATCTTT 15760

QY      601  gtaactacagaagtcacaaata  621
Db 15761 GTAGCTACAGAGTCACAATAA 15781
```

```
RESULT      8
US-09-611-529-7449/c
; Sequence 7449, Application US/09611529
; GENERAL INFORMATION:
; APPLICANT: George H. Shimer, Jr.
; APPLICANT: George H. Miller
; APPLICANT: Roberta S. Hare
; APPLICANT: Karen J. Shaw
; TITLE OR INVENTION: Staphylococcus aureus Related Compositions and Methods
; FILE REFERENCE: 1034/1C9630S1
; CURRENT APPLICATION NUMBER: US/09/611,529
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: US 09/417,811
; PRIOR FILING DATE: 1999-10-14
; PRIOR APPLICATION NUMBER: US 09/353,718
; PRIOR FILING DATE: 1999-07-14
; PRIOR APPLICATION NUMBER: US 09/266,557
; PRIOR FILING DATE: 1999-03-11
; PRIOR APPLICATION NUMBER: US 09/266,556
; PRIOR FILING DATE: 1999-03-11
; PRIOR APPLICATION NUMBER: US 09/266,555
; PRIOR FILING DATE: 1999-03-11
; PRIOR APPLICATION NUMBER: US 09/266,542
; PRIOR FILING DATE: 1999-03-11
; PRIOR APPLICATION NUMBER: US 09/266,541
; PRIOR FILING DATE: 1999-03-11
; PRIOR APPLICATION NUMBER: US 09/037,934
; PRIOR FILING DATE: 1998-03-10
; PRIOR APPLICATION NUMBER: US 09/036,720
; PRIOR FILING DATE: 1998-03-06
; PRIOR APPLICATION NUMBER: US 09/036,338
; PRIOR FILING DATE: 1998-03-06
; PRIOR APPLICATION NUMBER: US 09/036,334
; PRIOR FILING DATE: 1998-03-06
; PRIOR APPLICATION NUMBER: US 09/036,221
; PRIOR FILING DATE: 1998-03-06
; PRIOR APPLICATION NUMBER: US 09/036,137
; PRIOR FILING DATE: 1998-03-06
; PRIOR APPLICATION NUMBER: US 09/036,082
; PRIOR FILING DATE: 1998-03-06
; PRIOR APPLICATION NUMBER: US 09/036,081
; PRIOR FILING DATE: 1998-03-06
; PRIOR APPLICATION NUMBER: US 09/036,079
; PRIOR FILING DATE: 1998-03-06
; PRIOR APPLICATION NUMBER: US 09/035,913
; PRIOR FILING DATE: 1998-03-06
; PRIOR APPLICATION NUMBER: US 09/035,744
; PRIOR FILING DATE: 1998-03-06
; PRIOR APPLICATION NUMBER: US 08/827,356
; PRIOR FILING DATE: 1997-04-01
; PRIOR APPLICATION NUMBER: US 08/831,156
; PRIOR FILING DATE: 1997-04-01
; PRIOR APPLICATION NUMBER: US 60/014,477
; PRIOR FILING DATE: 1996-04-01
; PRIOR APPLICATION NUMBER: US 60/016,743
; PRIOR FILING DATE: 1996-05-02
; PRIOR APPLICATION NUMBER: US 60/020,016
; PRIOR FILING DATE: 1996-06-14
; NUMBER OF SEQ ID NOS: 7451
; SEQ ID NO 7449
; LENGTH: 812842
; TYPE: DNA
; ORGANISM: Staphylococcus aureus
US-09-611-529-7449
```



```

Query Match 98.5%; Score 611.4; DB 22; Length 812842;
Best Local Similarity 99.0%; Pred. No. 1.5e-104;
Matches 615; Conservative 0; Mismatches 6; Indels 0; Gaps 0

QY 1 atgaaanaaaatggacaacatcgatlaataatgaacaacgcgtggtgltacttaccatgtgca 60
Db 362980 ATGAAANAATGACAAATCGATTAAATGACAAATCGCGTGTAGTACTTACTAGTGCCA 362921

QY 61 gcatattglttgcctaaaccacatatcgataattatcttccagaataagataatgaa 120
Db 362920 GCATTTGTGTGGCTAAACCCACATATCGATTAATTTATCTTCACGATTAATAATGAAGTTGAA 362861

QY 121 aagatgacaacaatgataaanaatgtaaaagaacagcgagtaaaagataaaagcagca 180
Db 362860 AAGATTGGAACATATGATATAAATGTTAAAGAACACGCGCATGTAAGACATATAAGCAGCA 362801

QY 181 gctaaacctcaaatctcgagaagataaatctgaaagtggcagcgtatataatgaaatlccaagat 240
Db 362800 GCTAAACCTCAAAATTCGCAAAATATAAATCAAAAGTGGCAGCCTATATATGAATATCCAGAT 362741

QY 241 gctgattataaagaacacgatatatccagagcagcaaacacctgaacaaatlaaatgagat 300
Db 362740 GCTGATTATAAAGAACCGTATATTCACGAGCACGACAACCTGAAACAAATTAATAGAGT 362681

QY 301 gtaagctcttcgagaanaaatgtaactcactaga tgaatcaaaatattcgaatctcagagacac 360
Db 362680 GTAACCTTTGCGAGAGAANAATTAATACACTAGATGATCAAAATATTTCAATTGCAGAGACAC 362621

QY 361 acttcattgacgcgtccgacactcataatttcaaaatcttaaaagcagccaaaaaaggtaat 420
Db 362620 ACTTTCATTGTGACGCTCCGACACTATCAATTAATTAACAAATCTTAAAGCACACCAAAAAAGGTACT 362561

QY 421 atggtgtaaccttaaaagttggtatgaataaacacgtaagataaaatgacaagtataagat 480
Db 362560 ATGGTGTACTTTAAAGTTGGTATATGAACACGTAAGTATATAAATGCAAGTATAAGAGAT 362501

QY 481 gttaaagctacagatgtaggagttctagatgtaacaaaagtgaaagataaacaattaca 540
Db 362500 GTTAAAGCCAACAGATGTAGAACTTCTAGATGTAACAAAAGGTAAAGATTAACATTTAACAA 362441

QY 541 ttaattactctgagatgattacaatgaaagacagcgcttgggaaanaacgtataacattt 600
Db 362440 TTAATTACTTGTGATGATTACATATGAANAAGACAGCGCTTTGGGAAAACGTAATAATCTTT 362381

QY 601 gttagctacagaatgcaataaa 621
Db 362380 GTAGCTACAGAAAGTCAAAATTA 362360

RESULT 9
US-60-038-697-735
: Sequence 735, Application US/60038697
: GENERAL INFORMATION:
: APPLICANT: Lagace, Robert E.
: APPLICANT: Corley, Neil C.
: APPLICANT: Russo, Frank D.
: TITLE OF INVENTION: NOCLEFTIDE SEQUENCES OF STAPHYLOCOCCUS AUREUS
: NUMBER OF SEQUENCES: 1027
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: INCYTE PHARMACEUTICALS, INC.
: STREET: 3174 PORTER DRIVE
: CITY: PALO ALTO
: STATE: CALIFORNIA
: COUNTRY: USA
: ZIP: 94304
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: word Perfect 6.1 for Windows/MS-DOS 6.2

```

```

1  CURRENT APPLICATION DATA:
2  APPLICATION NUMBER: US/60/038,697
3  FILING DATE:
4  CLASSIFICATION:
5  ATTORNEY/AGENT INFORMATION:
6  NAME: CERRONE, MICHAEL C.
7  REGISTRATION NUMBER: 39,132
8  REFERENCE/DOCKET NUMBER: PQ-0001 US
9  TELECOMMUNICATION INFORMATION:
10 TELEPHONE: (415) 855-0555
11 TELEFAX: (415) 845-4166
12 INFORMATION FOR SEQ ID NO: 735:
13 SEQUENCE CHARACTERISTICS:
14 LENGTH: 45487 base pairs
15 TYPE: nucleic acid
16 STRANDEDNESS: single
17 TOPOLOGY: linear
18 MOLECULE TYPE: cDNA
19 IMMEDIATE SOURCE:
20 CLONE: SAC0735
21 US-60-038-697-735
22
23 Query Match 92.1%; Score 572.2; DB 28; Length 45487;
24 Best Local Similarity 99.5%; Pred. No. 2.3e-97; Mismatches 3; Indels 0; Gaps 0;
25 Matches 574; Conservative
26
27 1 atgaaataatgacaaatcgatataatgacaatcgctgtgtgtgtaactatcctagtgcga 60
28 44911 ATGAAAAATGACAAATCGATTATATGCAATCGCTGCTGTGTGTACTTATCTTAGTGCA 44970
29
30 61 gcaatttctgtcttaaccacataatcgataaattacttcacgataaaagataatgaa 120
31 44971 GCATTTTGTGTTGCTAAACCAATATCGATATATTTCTTCACGATAAAGATAAAGTAA 45030
32
33 121 aagattggaacaatgatgaataaattgaagaagacggcgagatgaagaataaaagacgaa 180
34 45031 AAGATTGGAACATGATGAATAAATGTTAAAGACAGCGGAGTAAAGATAAAGACGACAA 45090
35
36 181 gctaaacctcaaatctcgaaagaataatcgaagaatgacgataatcttgaaattccagat 240
37 45091 GCTAAACCTCAAAATCCGAAAGATTAATCGAAGTGGCAGGCTATATTGAAATTCAGAT 45150
38
39 241 gctgattataaagaaccgatatatccaggaccagaacaaccctgaaccaattaaatgaagt 300
40 45151 GCTGATTATTAAGAACCAGTATATTCAGAGACCGAACAACCTGAACATTTAATAGAGGT 45210
41
42 301 gtaagcttgcagaagaataatgaatcactatagaatgaatcaaatatttcaatttcagagac 360
43 45211 GTAAGCTTGCAGAAAGAAATGAATCACTAGATGATCAAAATATTTCAATTTCAGAGAC 45270
44
45 361 acctcaattgacgcgtcggaactataatcaatttcaaaactttaaagcagccaaaaaagtgagt 420
46 45271 ACTTTCATTGACCGCTCCGACACTATCAATTTTCAAAATCTTAAAGCGACCAAAAAAGGTG 45330
47
48 421 atgggtaccttaaaagtttggtaataaacaacgtaagttaaaatbaacaagtaataagaagt 480
49 45331 ATGGGTACTTTAAAGTTGGTAATGAACACGTAAGTATAAATATACCAAGTTAAGAGAT 45390
50
51 481 gctaaagctacagatgtagagagttctatagatgacacaaagaagttaagaataaacaatttaaca 540
52 45391 GTTAAGCCTACAGATGTAAGAGTTCTTGATGTAACAAAAAGGTAAAGATTAACAAATTAA 45450
53
54 541 ttaattacttgtatgattacaaatgaagaagcagcg 577
55 45451 TTAATTACTTGTGATGATTACATGAATAAAGACAGCGC 45487
56
57 RESULT 10
58 US-60-046-714-734
59 Sequence 734, Application US/60046714
60 GENERAL INFORMATION:
61 APPLICANT: Lagace, Robert E.
62

```


OY	1	atgaaataatggagcaaatgcattatgcaaatcgctggtgtgtacttacttccctagtgcga	60
Db	1995	ATGAAAGCAGTGGTGAATAGATTAAATACCTTTAATAGGCGTATTGTATATATTTTAGCT	2054
OY	61	gcatacttgcttgcataaccacatcgcataatctatcttcacgataaagaataaagatgcga	120
Db	2055	ATTATATTATTCGAAAGCCATATATTCGATTAATTTATCTACATGAAAAGATTAACGATCAT	2114
OY	121	aagcttgacaatatgataaaaaatgtaaaaagaacggcgagataaagataaagaacgcga	180
Db	2115	AAAAATTGAAATATATGATATAAAAAGGAAAAAGAACAGACAAACATCTTA-----AA	2165
OY	181	gctaaacctcaaatccgaaagataaactcgaaagtgagcagctatattgaaatccagat	240
Db	2166	TGCGAGCGCAAGATACCTTCCGATTAATCTMAATATGGCTGGTTATATAGAACTTCAGAC	2225
OY	241	gctgatatlaaagaacccaglatatccagagcagcaaacacctgcgaacaaatlaatagaagt	300
Db	2226	GCACAAATATAAAGAACACAGATATACCTCGGTCCGACAGACACAGAACATCAATATAGAGGT	2285
OY	301	gtaagcttcbaagaagaataatgaatacagtagtgatgcataaatcttaatttcagtcggagac	360
Db	2286	GTAAGTTTGCAGAGGTGATGATCTTTAATCAACGAAATATTTCATATGCTGGTGCAT	2345
OY	361	aacttcatltagccgtccgaactatcaatttacaatctttaaagcagcagcaaaaagtagt	420
Db	2346	ACGTTTACAGATCGTTGCGACATCAATTTACAAATTTMAATTAATCAGCCAAATGCGTAGT	2405
OY	421	atggtgtactttaaagtgtagtaagaacacgtaagatataaatagcagataataagat	480
Db	2406	AAAGCTGATTTTTTAAAACGTGGAATAATCAAAAGTATAAANATTAACCTAATAATACGTGAT	2465
OY	481	gttaagcttacaagatgcagagagctcagtagtgacaanaaagtgaaagatataaacaattaca	540
Db	2466	GTTAAGCCTACAGAGGTTAAGGCTTTAGACGAACATCCTAATTAAGAAAATTCATTAAACA	2525
OY	541	ttaatacttgtagatgattacaatigaanaagacagcgctttggtgaat-aacgtlaaatctt	599
Db	2526	TTAATTTACTTGCGCATGCTACCAACGAAGAAGACGGGTGTTGGAAACCAAGAAATATTT	2585
OY	600	tgtagctacagaagtaaatata 621	
Db	2586	CATAGCTACACAAATTTAAGCTAA 2607	

THIS PAGE BLANK (USPTO)

PE 07-JAN-1997; 97EP-0100117.
XX
PR 05-JAN-1996; 96US-0009861.
XX
(HUMA-) HUMAN GENOME SCI INC.
XX
PA
XX
Barash SC, Choi GH, Dillon PJ, Faanon MR, Kunsch CA,
PI
PI Rosen CA;
XX
XX
DR WPI; 1997-374922/35.

XX anti-S.aureus vaccines **page 1473**
PT stored on computer readable medium and used in the production of
PT Polynucleotide(s) and proteins derived from *Staphylococcus aureus*
PS Claim 1, Page 1472-1474; 3271pp; English.

page 1473
4; 3271pp; Eng

This sequence represents one of 5191 *Staphylococcus aureus* DNA sequences of the invention. The DNA sequences are recorded on a computer readable medium, preferably selected from a floppy or hard disk, random access memory (RAM), read-only memory (ROM) or CD-ROM. Homology searches using the *S. aureus* DNA sequences allows putative functions to be assigned so that protein-encoding or regulatory regions of commercial, therapeutic or industrial importance can be obtained. Specifically, sequences which are likely to encode antigens have been identified and these polypeptides can be used in a vaccine composition against *S. aureus* infection. The polypeptides can also be used in a kit for the immunodetection of *S. aureus* in a sample. *S. aureus* is implicated in numerous human diseases, including cellulitis, eyelid infections, food poisoning, osteomyelitis, skin and surgical wound infections, scalded skin syndrome, toxic shock syndrome, etc. Organisms transformed with the DNA sequences can be used for recombinant production of the polypeptides. The new DNA sequences (and their fragments) are useful as primers or probes for isolating homologues of any of the *S. aureus* DNA sequences contained on the computer readable medium.

Sequence 3733 BP; 1084 A; 718 C; 543 G; 1266 T; 122 other;

Query Match 98.4%; Score 611; DB 18; Length 3733;
 Best Local Similarity 100.0%; Pred. NO: 1.2e-118;
 Matches 611; Conservative 0; Indels 0; Gaps 0;

Qy	11	ggacaaatgattaaagacacgtgcgtggtgtrtacttaccctgaatggacgacatattgt	70
Db	720	ggacaaatgcatttaattacacattccgtctgctgtgtgacttattcttgaatggcagcattttgt	661
Qy	71	ttgtcaaccacacalcatgataattatcttcacgaataagaataagaatgaaagattgaac	130
Db	660	ttgttaataaccacattgtcgattattattcttcacgataaagattaaagattgaatttcac	601
Qy	131	aatagtaaaaatgtaaaagaacagcgagtaaaagataaaagcagcaactaaactc	190
Db	600	aattgtattaaaaaagtgtaaaaagacagcgagtaaaagattaaagcagcaactaaactc	541
Qy	191	aaatttcgaaagttaaatgcgaagtggcagcctatattgaattccagaatgcgatatta	250
Db	540	aaatttcgaaagattaaatgcgaagtggcagcctattgttaattccgaatgcgatattta	481
Qy	251	aagaaccagatataccagagccagcaaccctgaacaattaaataagagtgtaacttg	310
Db	480	aagaccagatatttccaggaccacgaacacctgtgacatttaattagagtgtaacttg	421
Qy	311	cagaagaanaatgatacattagatgatcaaaataattcaattgcagacaaacttcattg	370
Db	420	cagaagaanaatgatacactagatgatcaaaattttcaattgcagacaaacttcattg	361
Qy	371	accgtccgaactcaatttcaacaatcttaagcagcacaanaaagatgattgttact	430
Db	360	accgtccgaacttcaatttcaacaatttttaaacagccaaaaaaggatgattgttact	301
Qy	431	tttaagcttgtaatgaaacgtaagatataaaatgaaagtataagagattgtaagccta	490

QY	Db	QY	Db
300	TTAAGTTGGTAATGAACACAGCTAAGTATATAAATGACAACTATTAGAGATGTTAAGCTA	241	
491	CAGATGTAGgaagctctctcagatgaaacaaaagttaagaataacaacttaacattractt	550	
240	CAGATGTAGGAGCTTCTCGATGTAACAAAAAGGTAAGTAAAGTAAACAATTAATTAATCTT	181	
551	gtgagtattacaatgaaaaaacagcgcttgggaaaaagtaaatcttgtagctaacag	610	
180	GTGATGATTACAAATCAAAAAGCACAGCGCTTGGGAAAAACGTAAATATCTTGTAGCTACAG	121	
611	aagctcaaataa	621	
120	AAGTCAAAATAA	110	

RESULT 2

ID	T05868	standard; DNA; 3399 BP
1	1	1
2	2	2
3	3	3
4	4	4
5	5	5
6	6	6
7	7	7
8	8	8
9	9	9
10	10	10
11	11	11
12	12	12
13	13	13
14	14	14
15	15	15
16	16	16
17	17	17
18	18	18
19	19	19
20	20	20
21	21	21
22	22	22
23	23	23
24	24	24
25	25	25
26	26	26
27	27	27
28	28	28
29	29	29
30	30	30
31	31	31
32	32	32
33	33	33
34	34	34
35	35	35
36	36	36
37	37	37
38	38	38
39	39	39
40	40	40
41	41	41
42	42	42
43	43	43
44	44	44
45	45	45
46	46	46
47	47	47
48	48	48
49	49	49
50	50	50
51	51	51
52	52	52
53	53	53
54	54	54
55	55	55
56	56	56
57	57	57
58	58	58
59	59	59
60	60	60
61	61	61
62	62	62
63	63	63
64	64	64
65	65	65
66	66	66
67	67	67
68	68	68
69	69	69
70	70	70
71	71	71
72	72	72
73	73	73
74	74	74
75	75	75
76	76	76
77	77	77
78	78	78
79	79	79
80	80	80
81	81	81
82	82	82
83	83	83
84	84	84
85	85	85
86	86	86
87	87	87
88	88	88
89	89	89
90	90	90
91	91	91
92	92	92
93	93	93
94	94	94
95	95	95
96	96	96
97	97	97
98	98	98
99	99	99
100	100	100

AC T05868;

DT 14-AUG-1996 (first entry)

DE Chicken leucocytozoan DNA encoding immunogenic protein for vaccines.

Chicken leucocytozoan; immunogen; recombinant vaccine; protection;

05 Chicken leucocytozoan.

FH	Key	Location/Qualifiers
1	333	

```
FT      /*tag= a
```

ET /*tag= b

FT /note= "fragment referred to in the claims, for
 .FT use as insert in a recombinant vaccine
 FT against chicken leucocytozoan disease"

PN JP07284392-A

PD 31-OCT-1995.

PF 19-APR-1994; 94JP-0080643.

PR 19-APR-1994; 94JP-0080643.

PA (DOBU-) DOBUTSUYO SEIBUTSUGAKUTEKI SEIZAI KYOKAI.

XX

DR P-PSDB; R97866.

PT Chicken leucocytozoan immunogenic protein -

XX

XX

CC or a fragment of it can be used in a recombinant vaccine to immunise

CC and operatively linked to an expression regulatory sequence as in

XX

SQ Sequence 3399 BP; 1577 A; 508 C; 798 G; 516 T; 0 other;

[illegible]

Db 2069 aagaaagaagaagtaacacatgagaataagaaagaagcatgagaatc 2128
 Qy 161 gtaagataaaagcagcagcctaaacccaattccgaagaataatgaaagtcag 220
 Db 2129 atgaagaagaagaagaagtaacacatgagaagaagaagaagaagcctgaaag 2188
 Qy 221 gctatattgaaatccagatctgatataaagaacagtatataccagaccagcaac 280
 Db 2189 taatcatgaagaagaagaagaagtaacacatgagaagaagaagaagaagtaaac 2248
 Qy 281 ctgaacaataatagagtgatgagcttgcaagaagaataatcactagatgaca 340
 Db 2249 atgaagaagaagaagaagaagtaacacatgagaagaagaagaagaagtaacatgag 2308
 Qy 341 atattcattcagagacacacttccattgaccgctccgacatacattacaatccta 400
 Db 2309 aagaaagaagaagaagaagtaacacatgagaagaagaagaagaagtaacacatgagaaga 2368
 Qy 401 aagcagccaaagaagtgatgctgctttaaagttgtaatagaacagtaagta 460
 Db 2369 aagaagaagtaacacatgagaagaagaagaagaagtaacacatgagaagaagaag 2428
 Qy 461 aatgacaagataagagatgttaagccctacagatgtagagttctagaatgacaaga 520
 Db 2429 aagtaacacatgagaagaagaagaagaagtaacacatgagaagaagaagaagtaacac 2488
 Qy 521 gtaagaataaacaataatcattacttgatgcatgatacattgaaagcagcgtt 580
 Db 2489 atgaagaagaagaagaagaagtaacacatgagaagaagaagaagaagtaacacatgagaag 2548
 Qy 581 gggaagaacgtaaaatcttctgagctacagaagtcata 620
 Db 2549 aagaagaagtaacacatgagaagaagaagaagaagtaacata 2588
 RESULT 3
 087587
 ID 087587 standard; DNA: 1686 BP.
 AC 087587;
 XX
 Dt 19-DEC-1995 (first entry)
 DE DNA encoding Leucocytozoan protozoa structural protein epitope.
 XX
 KW Leucocytozoan protozoa: structural protein; epitope; vaccine; fowl.
 XX
 OS Leucocytozoan protozoa sp.
 XX
 JN JP0708995-A.
 XX
 PD 04-APR-1995.
 XX
 PF 10-SEP-1993; 93JP-0226078.
 XX
 PR 10-SEP-1993; 93JP-0226078.
 XX
 PA (DOBU-) DOBUTSUYO SEIBUTSUKUTTEKI SEIZAI KYOKAI.
 XX
 PA (N1SS-) N1SSEIKEN KK.
 XX
 DR WP1; 1995-167252/22.
 DR P-PSDB; R70491.
 XX
 PT Immune inducing polypeptide against Leucocytozoan protozoa - useful
 PT in production of vaccines for treatment of Leucocytozanosis in
 PT fowl.
 XX
 PS Claim 1; Page 12-14; 20pp; Japanese.
 CC 087587-89 encode polypeptides having a whole or partial epitope of a
 CC structural protein of Leucocytozoan protozoa (see K70491-93). The
 CC polypeptides and DNA encoding them are useful in the production of

CC vaccines for the treatment of Leucocytozanosis of fowl.
 XX
 SQ Sequence 1686 BP; 915 A; 129 C; 382 G; 260 T; 0 other;
 Query Match 10.8%; Score 66.8; DB 16; Length 1686;
 Best Local Similarity 45.2%; Pred. No. 1.1e-05;
 Matches 245; Conservative 0; Mismatches 297; Indels 0; Gaps 0;
 Qy 77 aaccacatctgaatttcttcacgataaagaatgaagaatgaaatgacaatag 136
 Db 188 aagaacaagatgagaagaagaagaagaaggaatgagaagaagaagaagaagaag 247
 Qy 137 ataaaatgtaaagaacagcgcgagtaagaatloaaagcagcagcctaaacccaattc 196
 Db 248 aagaagaagaagaagaagaagaagaagaagaagaagaagaagaagaagaagaagcagag 307
 Qy 197 cgaagaataatcgaagaagtcgacgctataatgaaatccagaatgctgatatgaagac 256
 Db 308 aagaagaagaagaagaagaagaagaagaagaagaagaagaagaagaagaagaagcagag 367
 Qy 257 cagtatatccagagacagcaaacaccgaaacaaatlaaagagtgatgagcctgcaag 316
 Db 368 aagaagaagaagaagaagaagaagaagaagaagaagaagaagaagaagaagaagcagag 427
 Qy 317 aaatgaatcactagaatgataaataatcattcattgagcaccacttccattgaccgtc 376
 Db 428 aacaaatgaagaagaagaagaagaagaagaagaagaagaagaagaagaagaagaagaag 487
 Qy 377 cgaactatcaattacaaatcttaagcagccaaagaagtagtatgctgctttaag 436
 Db 488 aacaaatgaagaagaagaagaagaagaagaagaagaagaagaagaagaagaagaagaag 547
 Qy 437 ttggtatgaacacgtaagataaataatgacaagaatgaagaatgttaagcctacagatg 496
 Db 548 aagaacaagaagaagaatgatatgctgaaagaagaagaagaagaagaagaagaagaagaag 607
 Qy 497 taggaattctagaatgacaagaagaagaagaagaagaagaagaagaagaagaagaagaagaag 556
 Db 608 aag 667
 Qy 557 attacaatgaagaagcagcgtttgggaagaagaagaagaagaagaagaagaagaagaagaag 616
 Db 668 tagaaag 727
 Qy 617 aa 618
 Db 728 aa 729
 RESULT 4
 N60472
 ID N60472 standard; DNA: 4590 BP.
 AC N60472;
 XX
 Dt 24-AUG-1991 (first entry)
 DE Sequence encoding the ring-infected Erythrocyte Surface Antigen
 DE (RESA).
 XX
 KW Malaria vaccine; antigen; epitope; ss.
 XX
 OS Plasmodium falciparum.
 XX
 FH Key Location/Qualifiers
 FH exon 801..995
 FT /*tag= a
 FT 1199..4225
 FT exon /*tag= b
 XX
 PN W08601802-A.

PD 27-MAR-1986.
 XX
 PF 11-SEP-1985; 85WO-0006960.
 XX
 PR 11-SEP-1984; 84AU-0007067.
 PR 11-SEP-1984; 84AU-0007066.
 PR 10-SEP-1985; 85AU-0047326.
 XX
 PA (HALL-) HALL INST MED RES.
 XX
 PI Kemp DJ, Anders R, Coppel RL, Brown G, Saint RB, Cowman AF;
 DR WPI; 1986-094065/14.
 DR P-PSDB; P60569.
 XX
 PT DNA coding for Plasmodium falciparum antigens - expressing
 PT polypeptide(s) having antigenicity of RESA or FIRA antigens of P
 PT falciparum
 XX
 PS Claim 4; Fig 1; 55pp; English.
 CC The inventors claim a novel DNA molecule which comprises a
 CC nucleotide sequence corresp. to all or a portion of the base
 CC sequence coding RESA (N60472) or FIRA (N60473). RESA and FIRA have
 CC antigenicity suitable for providing protective immunity against
 CC Plasmodium falciparum malarial infections.
 XX
 SO Sequence 4590 BP; 1933 A; 437 C; 673 G; 1547 T; 0 other;

Query Match 10.7%; Score 66.4; DB 7; Length 4590;
 Best Local Similarity 47.9%; Pred. No. 1.6e-05;
 Matches 224; Conservative 0; Mismatches 241; Indels 3; Gaps 1;
 QY 81 acatatgataatattcttcacgataagaataagatgaagaattgaacatatgataa 140
 DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 QY 3766 acatgagctggaagaataatgttagacacatgagctggaagaataatgttaga 3825
 DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 QY 141 aatgataaagaacagcgagtaagaataaagaacagcaactaacctcaattccgaa 200
 DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 DB 3826 agaatgtgaagaataatgttagaagaataatgttagaagaataatgttaga 3885
 QY 201 agataaactgaagctgacagctatattga--attccagatgctgatatataagaacc 257
 DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 DB 3886 agtagaagaataatgttagaagaataatgttagaagaataatgttagaagaata 3945
 QY 258 agtatatcccgaccagcaacacactggaacattaaatagaagtggttaagcttgcagaaga 317
 DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 DB 3946 tgttgaagaataatgttgaagaataatgttgaagaataatgttgaagaataatgttgaagaata 4005
 QY 318 aatgaatcactagatgatacaaatatttcaattgcagacacacttcatttcgctgcc 377
 DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 DB 4006 tgatgaagaataatgttgaagaataatgttagaagaataatgttagaagaataatgtt 4065
 QY 378 gaactatcaatttacaatctttaaagcagccaaaaggttagtattggttactttaagt 437
 DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 DB 4066 tgaagaataatgttagaagaataatgttgaagaataatgttagaagaataatgttaga 4125
 QY 438 tggtaataagaacacgctgaataaataatgacacagatagaagatgttaagctcaagatgt 497
 DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 DB 4126 agaaataatgttagaagaataatgttgaagaataatgttgaagaataatgttgaagaataatgttaga 4185
 QY 498 aggaagttcagatgaacaaaaggttaagaataaacaattacattaat 545
 DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 DB 4186 agaaataatgttgaagaataacacaaatgaagaataatgaagaataaataaataa 4233

DT 25-JUN-1999 (first entry)
 XX
 DE Base sequence of the plasmid prx-ires-bsr.
 XX
 KW Compox virus; bsr; viral vector; expression; apoptosis; resistance;
 KW crma; bcl-2; bcl-xl; FLIP; survival; IAP; ILP; adenovirus; cancer;
 KW autoimmune disease; graft rejection reaction; inflammation;
 KW inflammatory disease; ss.
 XX
 OS Synthetic.
 OS Compox virus.
 XX
 PN WC99J13073-A2.
 XX
 PD 18-MAR-1999.
 XX
 PF 07-SEP-1998; 98WO-JP04010.
 XX
 PR 08-SEP-1997; 97JP-0259235.
 XX
 PA (RPRG-) RPR GENCELL ASIA PACIFIC INC.
 XX
 PI Hamada H;
 XX
 DR WPI; 1999-243728/20.
 XX
 PT New apoptosis-resistant virus-sensitive cell
 XX
 PS Example 1; Page 38-41; 51pp; English.

XX
 CC The present invention describes an apoptosis-resistant virus-sensitive
 CC cell line into which an apoptosis resistance gene has been introduced.
 CC The recombinant viruses generated are capable of expressing apoptosis-
 CC associated genes. These can then be used in a variety of diseases for
 CC which the induction of apoptosis by gene transfer, or where the
 CC inhibition of harmful apoptosis, is therapeutic. The recombinant viruses
 CC are useful as vectors for gene therapy which can be applied to cancer
 CC therapy for destroying cancer cells selectively, the treatment of
 CC autoimmune diseases and graft rejection reaction, and apoptosis induction
 CC therapy for inflammatory cells in inflammatory diseases. Prior arts have
 CC encountered the problem where if an adenovirus vector capable of
 CC expressing an apoptosis-associated gene is introduced into animal cells,
 CC the cells producing the virus will be destroyed because the period of
 CC time required to induce cell death by apoptosis is shorter than that
 CC required to replicate and produce the virus, resulting in failure to
 CC obtain a recombinant virus having the integrated apoptosis-associated
 CC gene. In this invention an apoptosis-resistant 293 cell line (having an
 CC problem. The present sequence represents the base sequence of the
 CC plasmid prx-ires-bsr, which contains the compox virus bsr gene, and
 CC is used in an example from the present invention.
 XX
 SO Sequence 6644 BP; 2166 A; 1573 C; 1424 G; 1481 T; 0 other;

Query Match 9.6%; Score 59.8; DB 20; Length 6644;
 Best Local Similarity 44.3%; Pred. No. 0.00038;
 Matches 244; Conservative 0; Mismatches 307; Indels 0; Gaps 0;
 QY 71 ttgctaaccacatataatattcttcacgataaagaataagatgaagaattgac 130
 DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 DB 3735 ttgcataaaaaaataaaaaaataaaaaaataaaaaaataaaaaaataaaaaa 3794
 QY 131 aatatgaataaataatgttaagaacagcgagtaagaataaagaacgaactaacctc 190
 DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 DB 3795 aaaaaaataaaaaaataaaaaaataaaaaaataaaaaaataaaaaaataaaaaa 3854
 QY 191 aattccgaagaataaataatgtgaagctgacgctatattgaattccagatgctgatat 250
 DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 DB 3855 aaaaaaataaaaaaataaaaaaataaaaaaataaaaaaataaaaaaataaaaaa 3914
 QY 251 aagaacagtatatccgagccagcaacacctgacaataatagaagtgtytaagcttgg 310
 DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

RESULT 5
 X33181
 ID X33181 standard. DNA; 6644 BP.
 XX
 AC X33181;
 XX

Db	3915	aa	.3978
Oy	311	cagaagaaalgaatcactgatgatacaaatcttcaattgcagacacactttcatg	370
Db	3975	aa	4034
Oy	371	accgtcggaactatcaatttaacaattttaagcagccaaaaggtagtatggttact	430
Db	4035	aa	4094
Oy	431	ttaaagttygaatgaaacgcgttaagtataaaatgaacagtaagaagatg-tgaagcta	490
Db	4095	aa	4155
Oy	491	cagatltgaagtgcttcgaatgtaaacaaaaggtaagatataacaatttaacatttact	550
Db	4155	aa	4214
Oy	551	gtgatgttacatgtaaaagacagcgctttgggaaaaacgttaaatcttgtactcacg	610
Db	4215	aa	4274
Oy	611	aagtcacataa 621	
Db	4275	aaaaaaaaaaaa 4285	

RESULT	6
X33182	
ID	X33182 standard; DNA; 7372 BP
XX	
AC	X33182;
XX	
DT	25-JUN-1999 (first entry)

DE	Base sequence of the plasmid pRc-Bcl-X1-bst.
XX	
KW	Compx virus; bst; viral vector; expression; apoptosis; resistance
KW	crna; bcl-2; bcl-x1; FLIP; survivin; IAP; ILP; adenovirus; cancer;
KW	autoimmune disease; graft rejection reaction; inflammation;
KW	inflammatory disease; ss.

OS	Synthetic.	
OS	Homo sapiens.	
XX		
XX		
PN	MO9913073-A2.	
XX		
PD	18-MAR-1999.	
XX		
PF	07-SEP-1998;	98MO-JP04010.
XX		
PR	08-SEP-1997;	97JP-0259235.
XX		
PA	(RPRG-) RPR GENCELL	ASIA PACIFIC INC

PI	Hamada H;
XX	
DR	WPI; 1999-243728/20

PT New apoptosis-resistant virus-sensitive cell
XX
PS Example 2; Page 41-45; 51pp; English.

CC The present invention describes an apoptosis-resistant virus-sensitive
CC cell line into which an apoptosis resistance gene has been introduced.
CC The recombinant viruses generated are capable of expressing apoptosis-
CC associated genes. These can then be used in a variety of diseases for
CC which the induction of apoptosis by gene transfer, or where the
CC inhibition of harmful apoptosis, is therapeutic. The recombinant viruses
CC are useful as vectors for gene therapy, which can be applied to cancer
CC therapy for destroying cancer cells selectively, the treatment of
CC autoimmune diseases and graft rejection reaction, and apoptosis induction
CC therapy for inflammatory cells in inflammatory diseases. Prior arts have

CC encountered the problem where if an adenovirus vector capable of
CC expressing an apoptosis-associated gene is introduced into animal cells
CC the cells producing the virus will be destroyed because the period of
CC time required to induce cell death by apoptosis is shorter than that
CC required to replicate and produce the virus, resulting in failure to
CC obtain a recombinant virus having the integrated apoptosis-associated
CC gene. In this invention an apoptosis-resistant 293 cell line (having an
CC apoptosis resistant gene introduced) is established and overcomes the
CC problem. The present sequence represents the base sequence of the
CC plasmid ppx-bcl-xl-*bsr*, which contains the human Bcl-xl gene, and
CC is used in an example from the present invention.

Sequence 7372 BP; 2353 A; 1749 C; 1649 G; 1621 T; 0 other;

Query Match	9.6%;	Score 59.8;	DB 20;	Length 7372;
Best Local Similarity	44.3%;	Pred. No. 0.00039;		
Matches 244;	Conservative	0;	Mismatches 307;	Indels 0;
				Gaps 0;

Oy	71	ttgttaaacacatcgcatacttccgataagaacaagatgaagttagc	130
Db	4463	ttgcataaaaaaaaaaaaaaaaaaaaaaaaaaaaaaa	4522

QY 131 aatatgataaaatgttaaagacacgcygagftraagataaaaagcagcaagcttaacttc 190
 ||| | ||||| ||||| ||| | ||||| ||| ||||
 Db 4523 aaaaaaaataaaaaaaataaaaaaaataaaaaaaataaaaaaaataaaaaaaataaaaaaa 4583

QY 191 aaattccgaaagataaatcgaagtgcagctatatttgaattccagaatgctgatatta 250
||| ||| ||| ||| | ||| ||| ||| |
Db 4583 aaaaaaaaaaaaaaaaabbaaaaaaaaaaaaaaaaaaaaaaaaaaaaaa 4642

QY 251 aaagaaccagtatatccaggaaccagcacacacctgaaacaattaatagaygtgttaagcttgy 310
|| || | | | | | | | | | | | | | | | |
Db 4643 aaaaaaataaaaaaaaataaaaaaaaataaaaaaaaataaaaaaaaataaaaaaaaataaaaaaaa 471/2

OY 311 cagaagaaatgaatcactagtatgatacaaatatttcaattgcaggacacaccttcattg 370
| | | | | | | | | | | | | | |
Db 4703 aa 4762

[illegible]

431 ttaagttgtaatgaacacgtaagtataaattgacaagtataagagatgttaagccta 490

QY 491 cagatgttagtgattctagatgacaaanaagyttaagaataacaattaacattactt 550
| | | | | | | | | | | | | | |
Ph 4883 aaaaaaaaaaaaaaaaaaaaaaaa | | | | | | | | | | | | | | |

QY 551 gtgatgattacaatgaagaacagcgcttgggaagaacgtaaatccttgtagtacag 610
| | | | |
DB 4843 ||||| ||||| | |

QY	611	aagccaataa	621
E003			E017

RESULT	7
X33180	
ID	X33180 standard; DNA; 7797 BP
XX	
AC	X33180;
XX	
DT	25-JUN-1999 (first entry)

DE Cowpox virus bsr full length gene sequence.
XX
KM Cowpox virus: bsr; vital vector: expression; apoptosis; resistance
KM crma; bcl-2; bcl-xl; FLIP; survivin; IAP; ILP; adenovirus; cancer;
KM autoimmune disease; graft rejection reaction; inflammation;
KM

CC plasmid pK-Bcl 2-i-HCD 25, which contains the human Bcl-2 gene, and
 CC is used in an example from the present invention.
 XX
 SQ Sequence 7996 BP; 2463 A; 2015 C; 1829 G; 1689 T; 0 other;

Query Match 9.6%; Score 59.8; DB 20; Length 7996;
 Best Local Similarity 44.3%; Pred. No. 0.00039;
 Matches 244; Conservative 0; Mismatches 307; Indels 0; Gaps 0;

```

QY 71 ttgcataaccacatcgtatcgtatcctccagataagaataagatgaagatggaac 130
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 5087 ttgcataaaaaaataaaaaaataaaaaaataaaaaaataaaaaaataaaaaa 5146
QY 131 aatatgataaaatgtaaaagaacagcgagtaagaataaaagcgaagctaaaccc 190
   || ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 5147 aaaaaaataaaaaaataaaaaaataaaaaaataaaaaaataaaaaaataaaaaa 5206
QY 191 aaattccgaagaataaattcgaaagtgccagctatattgaattccagatgctgata 250
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 5207 aaaaaaataaaaaaataaaaaaataaaaaaataaaaaaataaaaaaataaaaaa 5266
QY 251 aagaacagctatattccagacagcagccctgaacaattatagatggttaagcttg 310
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 5267 aaaaaaataaaaaaataaaaaaataaaaaaataaaaaaataaaaaaataaaaaa 5326
QY 311 cagaagaataatgaatcactagatgataataattcattcagtcagacacattcattg 370
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 5327 aaaaaaataaaaaaataaaaaaataaaaaaataaaaaaataaaaaaataaaaaa 5386
QY 371 accgctcgaactatcaattacaaatcctaaagcagccaaaaggtagtagtgctact 430
   || ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 5387 aaaaaaataaaaaaataaaaaaataaaaaaataaaaaaataaaaaaataaaaaa 5446
QY 431 ttaagtggtgaatgaacacgttaagataaaatgacaaatgaagatggttaagccta 490
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 5447 aaaaaaataaaaaaataaaaaaataaaaaaataaaaaaataaaaaaataaaaaa 5506
QY 491 cagatgtagaggtctctagatgaacaagaagtaagaataaacattacattactt 550
   || ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 5507 aaaaaaataaaaaaataaaaaaataaaaaaataaaaaaataaaaaaataaaaaa 5566
QY 551 gtgagtattacaatgaaagacagcgcttggtgaaacacgtlaaatcttctgtgctacg 610
   || ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 5567 aaaaaaataaaaaaataaaaaaataaaaaaataaaaaaataaaaaaataaaaaa 5626
QY 611 aagtcacaata 621
   || ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 5627 aaaaaaataaaaa 5637

```

RESULT 9
 003875
 ID 003875 standard; DNA; 3095 BP.
 XX
 AC 003875;
 XX
 DT 24-AUG-1990 (first entry)
 XX
 DE Sequence encoding carboxylic terminal part of native GLURP.
 XX
 KW Plasmodium falciparum; antigen; malaria; vaccine; GLURP:ss
 XX
 OS Plasmodium falciparum.
 XX
 FH Key Location/Qualifiers
 FT J..2352
 FT CDS /tag=a
 FT /product=GLURP
 XX
 PN W09022811-A.
 PD 22-MAR-1990.

```

XX 18-SEP-1989; 89WO-0000218.
PF
XX
XX 03-MAR-1989; 89US-0218885.
PR
XX 03-MAR-1989; 89DK-0005191.
XX
XX (STAT-) STATENS SERUMINST.
PA
XX Driegel M, Borre M, Jepsen S, Vuust J, Rieneck K, Wind A;
PI Jakobsen PH;
XX
XX WPI: 1990-115998/15.
DR P-PDB: R05804.
XX
XX Polypeptide(s) derived from Plasmidium falciparum antigen - used in
XX vaccines and in production of antibodies, for diagnosis and
XX therapy of malaria.
XX
XX Disclosure: Fig 7; 108pp; English.
XX
XX An open reading frame of 2349 bps extends from the 5' terminal end of the
XX insert to a "TAA" stop codon. It is longest ORF found in the sequence.
XX Sequence displays some of the characteristics of other malaria nucleic
XX acid sequences: tandemly repeated motifs, high AT content, and a high
XX content of codons for glutamate. Three major repetitive sequences are:
XX one motif from bp 34 to bp 156 is repeated from bp 289 to bp 411; another
XX motif from bp 477 to bp 521 is repeated from bp 522 to bp
XX 566 and from bp 567 to bp 611; a third motif from bp 1174 to bp 1233 is
XX repeated tandemly 11 times. This last repetitive region consists of 360bp
XX repeats differing only in 3 bases GAT coding for aspartate. This region
XX is flanked to the 5' terminal od a degenerated 60 bp repeat. GC content
XX of the coding part of the insert is on average 30%, and of the non-coding
XX 3' terminal 11%.
XX
SQ Sequence 3095 BP; 1443 A; 300 C; 491 G; 861 T; 0 other;

```

Query Match 9.1%; Score 56.6; DB 11; Length 3095;
 Best Local Similarity 44.0%; Pred. No. 0.0016;
 Matches 239; Conservative 0; Mismatches 304; Indels 0; Gaps 0;

```

QY 72 tgcataaccacatcgtatcgtatcctccagataagaataagatgaagatggaac 131
   || ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1260 tgaatatagtagaggttggaagaattctaccagaagatgataaaatgaaaggtcaca 1319
QY 132 atattgataaaatgtaaaagaacagcgagtaagaataaaagcgaagctaaacctca 191
   || ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1320 tgaatatagtagaggttggaagaattctaccagaagataaaatgaaaggtcacaatga 1379
QY 192 aattccgaagaataatcgagaagtggtgaggtctattgaattccagatgctgatttaa 251
   || ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1380 aatagtagaggttggaagaattctaccagaagataaaatgaaaggttgaacaatgaat 1439
QY 252 agaacagatataatccaggacgaacacactgacaaatlaaataagagtgtagcttgc 311
   || ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1440 agtagaagtgtaagaataattctaccagaagataaaatgaaaggttcaacaatgaatag 1499
QY 312 agagaagaatgtaacactagatgatacaaatattcaattgcagagcacacttcatga 371
   || ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1500 agaggttgtaagaataattctaccagaagataaaatgaaaggttcaacatgaaatagtag 1559
QY 372 ccgtccgaactatcaattcaaatcttaagaagcgaacaaaggttagtagtggttact 431
   || ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1560 agttgaagaataattctaccagaagataaaatgaaaggttcaacatgaaatagtagaggt 1619
QY 432 taaagttgtaatgaaacacgtaagatlaaataagacagataagagatcttaagctac 491
   || ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1620 tgaagaataattctaccagaagataaaatgaaaggttcaacatgaaatagtagaggt 1679
QY 492 agatgtagaggttctgaatgaaacaaaaggtaagaataaacattacattacttg 551
   || ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1680 tgaagaataattctaccagaagataaaatgaaaggttcaacatgaaatagtagaggttga 1739

```


PA (INSP) INST PASTEUR.

XX Daubersies P, Druilhé P;

DR WPI: 1997-065464/06.

XX Plasmodium falciparum poly:peptide(s) and related nucleic acids -
PT derived from the liver stage antigen-3, useful for malaria vaccine
PT prodn. and diagnosis

PS Claim 20; Fig 1A-E; 69pp; French.

XX This sequence corresponds to the genomic sequence encoding a Plasmodium
CC falciparum strain K1 pre-erythrocytic liver stage antigen-3 (LSA-3)
CC protein (W24790). The gene sequence was isolated by screening a
CC P. falciparum strain T9/96 library with the serum from a missionary
CC treated by prophylaxis (for strain T6/96 see F9101286). Of 20 clones
CC isolated, clone 7295 was used to screen a library generated from Thai
CC strain K1. One clone contained a 6.85 kb insert including the sequence
CC presented here. The gene organisation comprises a first exon, a short
CC 168 bp intron and a 5 kb second exon containing a 1.8 kb region encoding
CC 7 blocks of 4 amino acid repeats and a 3' hydrophobic region
CC corresponding to a glycosyl-phosphatidylinositol membrane anchoring
CC sequence. The invention relates to new polypeptides of at least 10
CC amino acids derived from the LSA-3 polypeptide with the exception of
CC peptides W24791-4. The LSA-3 peptides can be used to raise antibodies
CC and as vaccines for immunotherapy of malaria.

XX Sequence 6152 BP; 2725 A; 463 C; 1215 G; 1749 T; 0 other;

Query Match 9.1%; Score 56.6; DB 18; Length 6152;

Best Local Similarity 47.2%; Pred. No. 0.0018;

Matches 241; Conservative 0; Mismatches 264; Indels 6; Gaps 2;

QY 104 ataagataagaatgaaagattgaaacaatcgtataaataatgtaagaacagcgagta 163
DB 4716 aagagcctaagtggaagaagtattataaagaagtgtaagaagaccagaataa 4775
QY 164 aagataaagaacagcagctaaacctcaattccgaagaataaaltcgaaggg---cag 220
DB 4776 aatacaaaaaaagaagaagtgatgataatgaagaagtaaaccaagaatgataag 4835
QY 221 gctatattgaattccagatgctgatatataaagaacagtatataccagaccagacac 280
DB 4836 tagaagtgtaaatgaaagtgaaagatatagaagaagatgtaaagaagatatagaaga 4895
QY 281 ctgaacaattaaatagaggtgtgaagcttgcagaagaataatgaatcagtagatcaaa 340
DB 4896 atataagaagaagaatgaagtgatagatatagaatagatatagatagtggaag 4955
QY 341 atattcaattcagacagacactttcattgacgcgtccgacatcataattacaactta 400
DB 4956 acaaa---agatgaagtataagatttaattagtcacaaaagaagaacgattgaaaag 5012
QY 401 aagcagccaaaaaagtagtactgctgataaagttgtaagtaaacagtaagata 460
DB 5013 aagcgaataaataaataaataaagaagtggaagaagtggttggtcttaaaaaaac 5072
QY 461 aatgacaagataagaagatgttaagcctacagatgtagagttcctagaatgaacaaaag 520
DB 5073 acgtagacgaagtaagtaaatgttcaaaaaatgtataagaagtgataagaagtat 5132
QY 521 gtaagaataacaataaataaataaataaataaagaagtgatgataagaacagcggtt 580
DB 5133 cttaagctttagatcaaaaaatgagttaactaatgttttaaaaaacaaatcagaattt 5192
QY 581 gggaaaaacgtaaatctttagtactacaga 611
DB 5193 ttgataaagttataaaacttcgtaaaaaata 5223

RESULT 12

T41852
ID T41852 standard; DNA: 9789 BP.
XX AC T41852;
XX 20-FEB-1997 (first entry)
DE cDNA encoding Plasmodium falciparum erythrocyte membrane protein.
XX Plasmodium falciparum; erythrocyte membrane protein; malaria;
XX detection; identification; treatment; prevention; parasite; ss.
XX OS Plasmodium falciparum MC type.
XX FH Key location/Qualifiers
XX CDS 326..9497
FT /tag= a
FT /product= Erythrocyte membrane protein
FT misc-feature 518..520
FT /tag= b
FT /transl_except= GTA encodes Tyrosine
FT 656..658
FT /tag= c
FT /transl_except= ATT encodes Leucine
FT 2909..2911
FT /tag= d
FT /transl_except= AAC encodes Aspartic acid
FT 3461..3463
FT /tag= e
FT /transl_except= GAA encodes Glutamine
FT 5546..5548
FT /tag= f
FT /transl_except= CCT encodes Arginine
FT 6254..6256
FT /tag= g
FT /transl_except= AAT encodes Lysine
FT 6257..6259
FT /tag= h
FT /transl_except= ATR encodes Tyrosine
FT 6263..6265
FT /tag= i
FT /transl_except= AAC encodes Lysine
FT 6269..6271
FT /tag= j
FT /transl_except= TTC encodes Isoleucine
FT 6272..6274
FT /tag= k
FT /transl_except= ATA encodes Histidine
FT 6275..6277
FT /tag= l
FT /transl_except= ATT encodes Asparagine
FT 6278..6280
FT /tag= m
FT /transl_except= GGA encodes Tryptophan
FT 7754..8478
FT /tag= n
PN WO9633736-A1.
PD 31-OCT-1996.
XX 26-APR-1996; 96WO-US05798.
PF 27-APR-1995; 95US-0430908.
XX (AFRY-) AFRYMAX TECHNOLOGIES NV.
XX Baruch DI, Howard RJ, Pasloske BL;
XX WPI: 1996-497376/49.
XX P-PSDB; W00384.
XX New Plasmodium falciparum erythrocyte membrane proteins - used to

FT			given in the specification for this DNA sequence
FT	misc_feature	10201..10260	
FT		/*tag- f	
FT		/note-	"these bases represent a line of missing text in the sequence listing in the specification. They are included to maintain the nucleotide numbering given in the specification for this DNA sequence"
FT	misc_feature	12001..12060	
FT		/*tag- g	
FT		/note-	"these bases represent a line of missing text in the sequence listing in the specification. They are included to maintain the nucleotide numbering given in the specification for this DNA sequence"
FT	misc_feature	13801..13860	
FT		/*tag- h	
FT		/note-	"these bases represent a line of missing text in the sequence listing in the specification. They are included to maintain the nucleotide numbering given in the specification for this DNA sequence"
FT	misc_feature	15601..15660	
FT		/*tag- i	
FT		/note-	"these bases represent a line of missing text in the sequence listing in the specification. They are included to maintain the nucleotide numbering given in the specification for this DNA sequence"
FT	misc_feature	17401..17460	
FT		/*tag- j	
FT		/note-	"these bases represent a line of missing text in the sequence listing in the specification. They are included to maintain the nucleotide numbering given in the specification for this DNA sequence"
FT	misc_feature	19201..19260	
FT		/*tag- k	
FT		/note-	"these bases represent a line of missing text in the sequence listing in the specification. They are included to maintain the nucleotide numbering given in the specification for this DNA sequence"
FT	misc_feature	21001..21060	
FT		/*tag- l	
FT		/note-	"these bases represent a line of missing text in the sequence listing in the specification. They are included to maintain the nucleotide numbering given in the specification for this DNA sequence"
FT	misc_feature	22801..22860	
FT		/*tag- m	
FT		/note-	"these bases represent a line of missing text in the sequence listing in the specification. They are included to maintain the nucleotide numbering given in the specification for this DNA sequence"
FT	misc_feature	24601..24660	
FT		/*tag- n	
FT		/note-	"these bases represent a line of missing text in the sequence listing in the specification. They are included to maintain the nucleotide numbering given in the specification for this DNA sequence"
FT	misc_feature	26401..26460	
FT		/*tag- o	
FT		/note-	"these bases represent a line of missing text in the sequence listing in the specification. They are included to maintain the nucleotide numbering given in the specification for this DNA sequence"
FT	misc_feature	28201..28260	
FT		/*tag- p	
FT		/note-	"these bases represent a line of missing text in the sequence listing in the specification. They are included to maintain the nucleotide numbering given in the specification for this DNA sequence"
FT	misc_feature	30001..30060	
FT		/*tag- q	
FT		/note-	"these bases represent a line of missing text in the sequence listing in the specification. They are included to maintain the nucleotide numbering given in the specification for this DNA sequence"

[illegible]

Db 19803 taaggacatttaaacatattacgaatgcaacaacgtacttgacttagaaggccaatcaalc 19862
Qy 524 aagataaacatataacttaattactgtgatgtatcaaatgaaagacagcggttg99 583
Db 19863 aagcgacgactgttgatgctgcttaatactgttaaaaaacaatgccaatacatatagacg9cg 19922
Qy 584 aaa 586
Db 19923 caa 19925

Search completed: March 14, 2001, 14:34:25
Job time: 25145 sec

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 2, 2001, 10:21:38 ; Search time 11.91 Seconds
(without alignments)
310.592 Million cell updates/sec

Title: US-09-292-437-3

Perfect score: 1076
Sequence: 1 MKRWTRNLTMTAGVILVLA.....YNEKGVWERKIFVATEVK 206

Scoring table:
BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 174772 seqs, 17957048 residues

Total number of hits satisfying chosen parameters: 174772

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :
1: /cgn2_6/ptodata/2/1aa/5A.COMB.pep:*
2: /cgn2_6/ptodata/2/1aa/5B.COMB.pep:*
3: /cgn2_6/ptodata/2/1aa/6.COMB.pep:*
4: /cgn2_6/ptodata/2/1aa/PCRTUS.COMB.pep:*
5: /cgn2_6/ptodata/2/1aa/Backfile1.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	137	12.7	223	US-08-961-083-162	Sequence 162, Appl
2	89.5	8.3	411	US-08-741-134-6	Sequence 6, Appl
3	88.5	8.2	289	US-08-961-083-20	Sequence 20, Appl
4	88.5	8.2	309	US-08-715-131-2	Sequence 2, Appl
5	82.5	7.7	783	US-08-922-837-2	Sequence 2, Appl
6	82	7.6	427	US-09-196-857-2	Sequence 2, Appl
7	82	7.6	558	US-08-285-440-6	Sequence 6, Appl
8	82	7.6	558	US-08-630-349-6	Sequence 6, Appl
9	80.5	7.5	759	US-08-676-967-1	Sequence 1, Appl
10	80.5	7.5	759	US-08-676-974-1	Sequence 1, Appl
11	80.5	7.5	759	US-09-098-487-1	Sequence 1, Appl
12	80	7.4	190	US-08-881-094-20	Sequence 20, Appl
13	80	7.4	257	US-08-715-204-6	Sequence 6, Appl
14	80	7.4	257	US-09-162-597-6	Sequence 6, Appl
15	80	7.4	273	US-08-137-175A-5	Sequence 5, Appl
16	80	7.4	273	US-08-479-017-5	Sequence 5, Appl
17	79.5	7.4	339	US-08-431-080-18	Sequence 18, Appl
18	79.5	7.4	339	US-08-938-534-18	Sequence 18, Appl
19	79.5	7.4	351	US-08-929-417-2	Sequence 2, Appl
20	79.5	7.4	765	US-08-663-112-2	Sequence 2, Appl
21	79	7.3	153	US-08-677-778B-1	Sequence 1, Appl
22	79	7.3	338	US-08-441-857-10	Sequence 10, Appl
23	79	7.3	338	US-08-193-159-10	Sequence 10, Appl
24	79	7.3	648	US-08-451-715A-4	Sequence 4, Appl
25	78	7.2	426	US-08-961-083-48	Sequence 48, Appl
26	78	7.2	581	US-08-961-083-132	Sequence 132, Appl
27	77	7.2	278	US-08-961-083-94	Sequence 94, Appl
28	77	7.2	451	US-08-435-454-4	Sequence 4, Appl

29	77	7.2	451	1	US-08-652-972A-4	Sequence 4, Appl
30	77	7.2	451	2	US-08-919-145-6	Sequence 6, Appl
31	77	7.2	451	3	US-08-870-126-4	Sequence 4, Appl
32	77	7.2	451	4	US-09-344-889-6	Sequence 6, Appl
33	77	7.2	451	4	PCR-US96-06231A-4	Sequence 4, Appl
34	77	7.2	571	3	US-08-961-083-4	Sequence 4, Appl
35	77	7.2	714	2	US-08-990-114-3	Sequence 3, Appl
36	76.5	7.1	331	2	US-08-828-242-3	Sequence 3, Appl
37	76.5	7.1	331	2	US-08-910-927B-5	Sequence 5, Appl
38	76.5	7.1	1964	2	US-08-790-912-3	Sequence 3, Appl
39	76.5	7.1	2052	2	US-08-790-912-2	Sequence 2, Appl
40	76	7.1	1664	2	US-08-642-846-2	Sequence 2, Appl
41	75.5	7.0	1674	2	US-08-968-542C-12	Sequence 12, Appl
42	75	7.0	593	1	US-08-296-362-2	Sequence 2, Appl
43	74.5	6.9	274	3	US-08-937-271-6	Sequence 6, Appl
44	74.5	6.9	294	1	US-08-137-175A-9	Sequence 9, Appl
45	74.5	6.9	294	3	US-08-479-017-9	Sequence 9, Appl

ALIGNMENTS

RESULT 1
US-08-961-083-162
Sequence 162, Application US/08961083
Patent No. 6159469
GENERAL INFORMATION:
APPLICANT: Choi et. al.
TITLE OF INVENTION: Streptococcus pneumoniae Antigenes and Vaccines
NUMBER OF SEQUENCES: 452
CORRESPONDENCE ADDRESS:
ADDRESSEE: Human Genome Sciences, Inc.
STREET: 9410 Key West Avenue
CITY: Rockville
STATE: Maryland
COUNTRY: USA
ZIP: 20850
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 3.50 inch, 1.4mb storage
COMPUTER: HP Vectra 486/33
OPERATING SYSTEM: MSDOS version 6.2
SOFTWARE: ASCII Text
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/961,083
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Brookes, A. Anders
REGISTRATION NUMBER: 36,373
REFERENCE/DOCKET NUMBER: PB340P2
TELECOMMUNICATION INFORMATION:
TELEPHONE: (301) 309-8512
TELEFAX: (301) 309-8504
INFORMATION FOR SEQ ID NO: 162:
SEQUENCE CHARACTERISTICS:
LENGTH: 223 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-961-083-162

Query Match 12.7%; Score 137; DB 3; Length 223;
Best local Similarity 30.8%; Pred. No. 8.8e-07;
Matches 45; Conservative 23; Mismatches 62; Indels 16; Gaps 7;
QY 64 QIRKDKSKVAGYIEFDADIKRVPYGPATPEQNLNGVFAEENE-SLDDQNTS--IAGH 120
DB 52 QVSDPDVAYGYLISPLSEIMEPVYLG-ADYHRLGMLAHVDGTPPLDQTCIRSYIAGH 110


```

? OPERATING SYSTEM: DOS
? SOFTWARE: FastSEO for Windows Version 2.0
? CURRENT APPLICATION DATA:
? APPLICATION NUMBER: US/08/922.837
? FILING DATE:
? CLASSIFICATION: 435
? PRIOR APPLICATION DATA:
? APPLICATION NUMBER:
? FILING DATE:
? ATTORNEY/AGENT INFORMATION:
? NAME: Dickinson, Q. Todd
? REGISTRATION NUMBER: 28,354
? REFERENCE/DOCKET NUMBER: GAI0077
? TELECOMMUNICATION INFORMATION:
? TELEPHONE: 215/994-2252
? TELEFAX: 215/994-2222
? TELEX:
? INFORMATION FOR SEQ ID NO: 2:
? SEQUENCE CHARACTERISTICS:
? LENGTH: 783 amino acids
? TYPE: amino acid
? STRANDEDNESS: single
? TOPOLOGY: linear
?
US-08-922-837-2
```

[illegible]

RESULT 6
US-09-196-857-2
Sequence 2, Application US/09196857A
Patent No. 6100069
GENERAL INFORMATION:
APPLICANT: zalcain, Magdalena
APPLICANT: Brown, James R.
APPLICANT: Biswas, Sanjoy
APPLICANT: Throup, John P.
APPLICANT: Lawlor, Elizabeth J.
APPLICANT: Mooney, Jeffrey
APPLICANT: Zhong, Yivi
APPLICANT: Debouck, Christine
APPLICANT: Jaworski, Deborah D.
APPLICANT: Wang, Min
APPLICANT: Warren, Richard L.
APPLICANT: Schilling, Lisa K.
TITLE OF INVENTION: No. 6100069e91 tig
FILE REFERENCE: GM10118
CURRENT APPLICATION NUMBER: US/09/196,857A
CURRENT FILING DATE: 1998-11-20
EARLIER APPLICATION NUMBER: 60/082,418

```

; EARLIER FILING DATE: 1997-11-20
; NUMBER OF SEQ ID NOS: 5
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 2
; LENGTH: 427
; TYPE: PRT
; ORGANISM: Streptococcus pneumoniae
; US-09-196-857-2

```

```

Query Match          7.6%: Score 82; DB 3; Length 427;
Best Local Similarity 21.8%: Pred. NO. 1.4;
Matches 43; Conservative 30; Mismatches 70; Indels 54; Gaps 8;

```

```

QY 18 LVAAYLFAPKPH-----IDNYLHDKDKDEKIEQYDKNVKQASKDKKQAKP 63
DB 105 VIAAEVYTKPEVKLGDKYKNLEVSVDVEKEVTDADVEERIEERNNTLTLYLKEAAEAGND 164
QY 64 QIPKSKSVAGYIEIPADIKEPYPCGPATPEQLNKGVSFAEENESLDDQNISAGH--- 120
DB 165 TVVID---FVGSINGVERD-----GKGGENFSLGISGQFIPEFDQ---LVGHSAG 210
QY 121 -----TFIDRPVYQFTNKAARKGSMVYFKVGNETPKYKMTSIRDYKPTPDVGYLDEOK 173
DB 211 ETVDVIVTEPE--DYQAEGL-AGKEAFV-----TTIHEVKAKEVPALDEL 254
QY 174 GKDKQLTITCDYNEK 190
DB 255 AKDIDEVEETLADLKEK 271

```

RESULT 7

```

US-08-285-440-6
; Sequence 6, Application US/08285440
; Patent No. 5532337

```

GENERAL INFORMATION:

```

APPLICANT: Ken Ichiro HAYASHI et al.

```

```

TITLE OF INVENTION: POLYPEPTIDE

```

```

NUMBER OF SEQUENCES: 32

```

```

CORRESPONDENCE ADDRESS:

```

```

ADDRESSEE: Wenderoth, Lind & Ponack

```

```

STREET: 805 Fifteenth Street, N.W., #700

```

```

CITY: Washington

```

```

STATE: D.C.

```

```

COUNTRY: U.S.A.

```

```

ZIP: 20005

```

```

COMPUTER READABLE FORM:

```

```

MEDIUM TYPE: Diskette, 5.25 inch, 500 kb

```

```

COMPUTER: IBM Compatible

```

```

OPERATING SYSTEM: MS-DOS

```

```

SOFTWARE: Displaywrite

```

```

CURRENT APPLICATION DATA:

```

```

APPLICATION NUMBER: US/08/285.440

```

```

FILING DATE:

```

```

CLASSIFICATION: 536

```

```

PRIOR APPLICATION DATA:

```

```

APPLICATION NUMBER: 07/858,947

```

```

FILING DATE: March 27, 1992

```

```

ATTORNEY/AGENT INFORMATION:

```

```

NAME: Warren M. Cheek, Jr.

```

```

REGISTRATION NUMBER: 33,367

```

```

REFERENCE/DOCKET NUMBER:

```

```

TELECOMMUNICATION INFORMATION:

```

```

TELEPHONE: 202-371-8850

```

```

TELEFAX:

```

```

TELEX:

```

```

INFORMATION FOR SEQ ID NO: 6:

```

```

SEQUENCE CHARACTERISTICS:

```

```

LENGTH: 558 amino acids

```

```

TYPE: amino acid

```

```

STRANDEDNESS: single

```

```

TOPOLOGY: linear

```

```

MOLECULE TYPE: peptide

```

```

HYPOTHETICAL:
ANTI-SENSE:
FRAGMENT TYPE:
ORIGINAL SOURCE:
ORGANISM:
STRAIN:
INDIVIDUAL ISOLATE:
DEVELOPMENTAL STAGE:
HAPLOTYPE:
TISSUE TYPE:
CELL LINE:
ORGANELLE:
IMMEDIATE SOURCE:
LIBRARY:
CLONE:
POSITION IN GENOME:
CHROMOSOME/SEGMENT:
MAP POSITION:
UNITS:
FEATURE:
NAME/KEY:
LOCATION:
IDENTIFICATION METHOD:
OTHER INFORMATION:
PUBLICATION INFORMATION:
AUTHORS:
TITLE:
JOURNAL:
VOLUME:
ISSUE:
PAGES:
DATE:
DOCUMENT NUMBER:
FILING DATE:
PUBLICATION DATE:
RELEVANT RESIDUES IN SEQ ID NO:

```

```

US-08-285-440-6

```

```

Query Match          7.6%: Score 82; DB 1; Length 558;
Best Local Similarity 25.9%: Pred. NO. 2;
Matches 29; Conservative 16; Mismatches 43; Indels 24; Gaps 4;

```

```

QY 30 DNYLHDKDKDEKIEQ-----YDKNVKQASDKKQAKPQIPKSKSVAGYIE 77
DB 176 ENKKEDKKEKEEBEKKPKRGSIGENQEEKTKYQAKREKLEQEDKPTKKE-----E 227
QY 78 IPDADIKEPYVPGPATPEQLNKGVSFAE---ENESLDDQNISAGHFTIDRP 126
DB 228 IKDEIKKKDKKEKKEVKSFMDRKKGFTEVKSGNGEFMTHKLKHTEMT-SRP 278

```

RESULT 8

```

US-08-630-349-6
; Sequence 6, Application US/08630349
; Patent No. 5739008

```

GENERAL INFORMATION:

```

APPLICANT: Ken Ichiro HAYASHI et al.

```

```

TITLE OF INVENTION: POLYPEPTIDE

```

```

NUMBER OF SEQUENCES: 32

```

```

CORRESPONDENCE ADDRESS:

```

```

ADDRESSEE: Wenderoth, Lind & Ponack

```

```

STREET: 805 Fifteenth Street, N.W., #700

```

```

CITY: Washington

```

```

STATE: D.C.

```

```

COUNTRY: U.S.A.

```

```

ZIP: 20005

```

```

COMPUTER READABLE FORM:

```

```

MEDIUM TYPE: Diskette, 3.5 inch, 144 mb

```

```

COMPUTER: IBM Compatible

```

```

OPERATING SYSTEM: MS-DOS

```

```

SOFTWARE: Wordperfect 5.1

```



```

CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/630,349
FILING DATE: April 10, 1996
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/285,440
FILING DATE: August 4, 1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 07/858,947
FILING DATE: March 27, 1992
ATTORNEY/AGENT INFORMATION:
NAME: Warren M. Cheek, Jr.
REGISTRATION NUMBER: 33,367
INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 558 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
US-08-630-349-6

```

```

Query Match          7.6%; Score 82; DB 1; Length 558;
Best Local Similarity 25.9%; Pred. No. 2;
Matches 29; Conservative 16; Mismatches 43; Indels 24; Gaps 4;

QY 30 DNVLHDKDEKIEQ-----YDKNVKQASKDKQOAKPOIPKDKSVAGIYE 77
DB 176 ENKKEKEKEEKEEKEPKKSGIGENGEKGYKQAKREKLODKPFFKE-----E 227
QY 78 IPDADIKPEVPGAPATPEQLNRCVSFAE--ENESLDDONISAGHTFIDRP 126
DB 228 IKDEKIKKKEPEEVEKSFMDRKKGFTEVKSQNGEPMTIKLKTENTF--SRP 278

```

```

RESULT 9
US-08-676-967-1
Sequence 1, Application US/08676967
Patent No. 5747317
GENERAL INFORMATION:
APPLICANT: COLLINS, KATHLEEN
TITLE OF INVENTION: Human Telomerase
NUMBER OF SEQUENCES: 10
CORRESPONDENCE ADDRESS:
ADDRESS: Science & Technology Law Group
STREET: 268 Bush Street, Suite 3200
CITY: San Francisco
STATE: CA
COUNTRY: USA
ZIP: 94104
COMPUTER READABLE FORM:
MEDIUM TYPE: floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/676,967
FILING DATE:
CLASSIFICATION: 530
ATTORNEY/AGENT INFORMATION:
NAME: Osman Ph.D., Richard A
REGISTRATION NUMBER: 36,627
REFERENCE/DOCKET NUMBER: UCB96-055
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415)343-4341
TELEFAX: (415)343-4342
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 759 amino acids
TYPE: amino acid
STRANDEDNESS:
TOPOLOGY: not relevant

```

```

MOLECULE TYPE: peptide
US-08-676-967-1
Query Match          7.5%; Score 80.5; DB 1; Length 759;
Best Local Similarity 21.6%; Pred. No. 4.5;
Matches 35; Conservative 32; Mismatches 80; Indels 15; Gaps 4;

QY 35 DKDKDEKIEQYDKNVKEQASKDKKQOAKPOIPKDKSVAGIETIPDADIKPEVPGPATP 94
DB 227 DMEEEENDDDDDDDEEDGDFDEDEEENI---ESVYTPVQIQKAVR---PAPAKS 280
QY 95 EQLNRCVSAEENESLDDONISAGHTFIDRPVQFTNLKAAGKSNVFRVGNETPKRYK 154
DB 281 SDHSEEDSDLEEDSDIDGELAQSPTSTEEQEDKAVQVSNKKR-----KLPSDVNEGK 335
QY 155 MTSIR---DVKPTDVGVLDEQKQKQQLTLITCDYNETG 192
DB 336 TVEIRNLSPDEEELGELLQDFGELKYRIVRIHPDTEHSGK 377

```

```

RESULT 10
US-08-676-974-1
Sequence 1, Application US/08676974
Patent No. 5770422
GENERAL INFORMATION:
APPLICANT: COLLINS, KATHLEEN
TITLE OF INVENTION: Human Telomerase
NUMBER OF SEQUENCES: 10
CORRESPONDENCE ADDRESS:
ADDRESS: Science & Technology Law Group
STREET: 268 Bush Street, Suite 3200
CITY: San Francisco
STATE: CA
COUNTRY: USA
ZIP: 94104
COMPUTER READABLE FORM:
MEDIUM TYPE: floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/676,974
FILING DATE:
CLASSIFICATION: 530
ATTORNEY/AGENT INFORMATION:
NAME: Osman Ph.D., Richard A
REGISTRATION NUMBER: 36,627
REFERENCE/DOCKET NUMBER: UCB96-055
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415)343-4341
TELEFAX: (415)343-4342
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 759 amino acids
TYPE: amino acid
STRANDEDNESS:
TOPOLOGY: not relevant
MOLECULE TYPE: peptide
US-08-676-974-1
Query Match          7.5%; Score 80.5; DB 1; Length 759;
Best Local Similarity 21.6%; Pred. No. 4.5;
Matches 35; Conservative 32; Mismatches 80; Indels 15; Gaps 4;

QY 35 DKDKDEKIEQYDKNVKEQASKDKKQOAKPOIPKDKSVAGIETIPDADIKPEVPGPATP 94
DB 227 DMEEEENDDDDDDDEEDGDFDEDEEENI---ESVYTPVQIQKAVR---PAPAKS 280
QY 95 EQLNRCVSAEENESLDDONISAGHTFIDRPVQFTNLKAAGKSNVFRVGNETPKRYK 154
DB 281 SDHSEEDSDLEEDSDIDGELAQSPTSTEEQEDKAVQVSNKKR-----KLPSDVNEGK 335

```

QY 155 MTSIR----DYKPTDVGVLDEOKGDKOLITITCDYNEKKG 192
Db 336 TVFIRNLSEEEBELGLLOQFGLKAYRVLVHPDTEHSGK 377

RESULT 11

US-09-098-487-1
; Sequence 1, Application US/09098487
; Patent No. 5917025
; GENERAL INFORMATION:
; APPLICANT: COLLINS, Kathleen
; TITLE OF INVENTION: Human Telomerase
; NUMBER OF SEQUENCES: 11
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Science & Technology Law Group
; STREET: 268 Bush Street, Suite 3200
; CITY: San Francisco
; STATE: CA
; COUNTRY: USA
; ZIP: 94104
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/098,487
; FILING DATE:
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Osman Ph.D., Richard A
; REGISTRATION NUMBER: 36,627
; REFERENCE/DOCKET NUMBER: UCB96-055
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415)343-4341
; TELEFAX: (415)343-4342
; INFORMATION FOR SEO ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 759 amino acids
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: not relevant
; MOLECULE TYPE: peptide
US-09-098-487-1

Query Match 7.5%; Score 80.5; DB 2; Length 759;
Best Local Similarity 21.6%; Pred. No. 4.5;
Matches 35; Conservative 32; Mismatches 80; Indels 15; Gaps 4;

QY 35 DKDKDEKIEGYDKNVKQKQAKPOIPKDKSVAGYIEIPDADIEKPYGYPATP 94
Db 227 DMEEEENDDDDDDEEGVDDDEEEENT---ESKVTKVOJOKRAVKR---PAPAKS 280
QY 95 EQLNRGVSAEENESLDQONISIGHFTIDRPNYQFTNLKAAGKSNVYFVGNETRRYK 154
Db 281 SDHSESDSDLEESIDGELAQSDSTEEQEDKAVGVSNKKR-----KLPSVNSGK 335
QY 155 MTSIR----DYKPTDVGVLDEOKGDKOLITITCDYNEKKG 192
Db 336 TVFIRNLSEEEBELGLLOQFGLKAYRVLVHPDTEHSGK 377

RESULT 12

US-08-881-094-20
; Sequence 20, Application US/08881094A
; Patent No. 6022739
; GENERAL INFORMATION:
; APPLICANT: Ryan, Clarence A
; APPLICANT: Pearce, Gregory L
; APPLICANT: McGurl, Barry F
; TITLE OF INVENTION: SystemIn

; FILE REFERENCE: 7555-000001CPB
; CURRENT APPLICATION NUMBER: US/08/881,094A
; CURRENT FILING DATE: 1997-07-09
; EARLIER APPLICATION NUMBER: 08/308,887
; EARLIER FILING DATE: 1994-09-19
; EARLIER APPLICATION NUMBER: PCT/US93/02428
; EARLIER FILING DATE: 1993-03-18
; EARLIER APPLICATION NUMBER: 07/885,412
; EARLIER FILING DATE: 1992-03-19
; EARLIER APPLICATION NUMBER: 07/528,956
; EARLIER FILING DATE: 1990-05-25
; EARLIER APPLICATION NUMBER: PCT/US91/03685
; EARLIER FILING DATE: 1991-05-24
; NUMBER OF SEO ID NOS: 43
; SOFTWARE: PatentIn Ver. 2.0
; SEO ID NO 20
; LENGTH: 190
; TYPE: PRT
; ORGANISM: Capsicum annuum
US-08-881-094-20

Query Match 7.4%; Score 80; DB 3; Length 190;
Best Local Similarity 21.6%; Pred. No. 0.69;
Matches 30; Conservative 23; Mismatches 52; Indels 34; Gaps 4;

QY 34 HDKDKDEKIEGYDKNVKQKQAKPOIPKDK-----SKVAGYIE 77
Db 23 HDKGEDEKIEKTEPDSODIGNKGDQAEKPKVREBSGDIEKIVENETSSQISIKIE 82
QY 78 IPDA-----DIKEPYPGPAPPEOLNRGVSAE-----NESLDQONISIGH 119
Db 83 EDDAQDNKIVEEDEDEKETEAKATPPQDINNKGDAQAEKPTVEHHEDEGDEKKEIKYKG 142
QY 120 HTFIDRPNYQFTNLKAARK 138
Db 143 DDALIEPKVEYEEEREKRR 161

RESULT 13

US-08-715-204-6
; Sequence 6, Application US/08715204
; Patent No. 5874286
; GENERAL INFORMATION:
; APPLICANT: Bandman, Olga
; APPLICANT: Au-Young, Janice
; APPLICANT: Goll, Surya K.
; APPLICANT: Hillman, Jennifer
; APPLICANT: Zweiger, Gary B.
; TITLE OF INVENTION: A NOVEL TUMOR PROTEIN
; NUMBER OF SEQUENCES: 7
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Incyte Pharmaceuticals, Inc.
; STREET: 3174 Porter Drive
; CITY: Palo Alto
; STATE: CA
; COUNTRY: U.S.
; ZIP: 94304
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq Version 1.5
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/715,204
; FILING DATE: Filed Herewith
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Billings, Lucy J.
; REGISTRATION NUMBER: 36,749
; REFERENCE/DOCKET NUMBER: PF-0126 US

TELECOMMUNICATION INFORMATION:
 TELEPHONE: 415-845-4166
 TELEFAX: 415-845-4166
 INFORMATION FOR SEQ ID NO: 6:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 257 amino acids
 TYPE: amino acid
 STRANDEDNESS: single
 TOPOLOGY: linear
 MOLECULE TYPE: peptide
 IMMEDIATE SOURCE:
 LIBRARY: Genbank
 CLONE: 1072344
 US-08-715-204-6

Query Match 7.4%; Score 80; DB 2; Length 257;
 Best Local Similarity 19.7%; Pred. No. 1.1;
 Matches 43; Conservative 37; Mismatches 86; Indels 52; Gaps 6;
 QY 35 DKDKDEKIEQYDKNNVEQASDKDKKQQA-----KQIIPKDSKVAGYIEIPDAD---- 82
 DB 5 NKKPKNEKLEKFAKELQSDSDDAVYIEQPTVEPKLPQNDSSSNKIVLSQAEKDLL 64
 QY 83 -----IKEPVYPGPATPEQNLNGVSPAEENESLDDQNISAGHTFIDR-PNY 128
 DB 65 RTELDKTEEEISTLKQVLSARQKHAELKRLGLFPFSELSQDINSKLTVDTDACHTF 124
 QY 129 QFTNLAAKAGSMVYFK-----VGNETR-----KYKMTSIRDYKPPIDGVY 168
 DB 125 IEINIOKKKKQSMYIKRLSKNIQTVPILTSEKRLILHAFLVILKKSSILKSL-----L 178
 QY 169 IDEQKGGDKQLTLITCDNYNEKTGVWEKRRKIFVATEVK 206
 DB 179 LMQOYOKTAEVAATSDTYKEKMNDRNSSLFSPFSK 216

RESULT 14
 US-09-162-597-6
 Sequence 6, Application US/09162597
 Patent No. 6043343
 GENERAL INFORMATION:
 APPLICANT: Bandman, Olga
 APPLICANT: Au-Young, Janice
 APPLICANT: Goli, Surya K.
 APPLICANT: Hillman, Jennifer
 APPLICANT: Zweiger, Gary B.
 TITLE OF INVENTION: A NOVEL TUMOR PROTEIN
 NUMBER OF SEQUENCES: 7
 CORRESPONDENCE ADDRESS:
 STREET: 3174 Porter Drive
 CITY: Palo Alto
 STATE: CA
 COUNTRY: U.S.
 ZIP: 94304
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Diskette
 COMPUTER: IBM Compatible
 OPERATING SYSTEM: DOS
 SOFTWARE: FASTSEQ Version 1.5
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/09/162,597
 FILING DATE:
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: 08/715,204
 FILING DATE:
 ATTORNEY/AGENT INFORMATION:
 NAME: Billings, Lucy J.
 REGISTRATION NUMBER: 36,749
 REFERENCE/DOCKET NUMBER: PF-0126 US
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: 415-855-0555

TELEFAX: 415-845-4166
 INFORMATION FOR SEQ ID NO: 6:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 257 amino acids
 TYPE: amino acid
 STRANDEDNESS: single
 TOPOLOGY: linear
 MOLECULE TYPE: peptide
 IMMEDIATE SOURCE:
 LIBRARY: Genbank
 CLONE: 1072344
 US-09-162-597-6

Query Match 7.4%; Score 80; DB 3; Length 257;
 Best Local Similarity 19.7%; Pred. No. 1.1;
 Matches 43; Conservative 37; Mismatches 86; Indels 52; Gaps 6;
 QY 35 DKDKDEKIEQYDKNNVEQASDKDKKQQA-----KQIIPKDSKVAGYIEIPDAD---- 82
 DB 5 NKKPKNEKLEKFAKELQSDSDDAVYIEQPTVEPKLPQNDSSSNKIVLSQAEKDLL 64
 QY 83 -----IKEPVYPGPATPEQNLNGVSPAEENESLDDQNISAGHTFIDR-PNY 128
 DB 65 RTELDKTEEEISTLKQVLSARQKHAELKRLGLFPFSELSQDINSKLTVDTDACHTF 124
 QY 129 QFTNLAAKAGSMVYFK-----VGNETR-----KYKMTSIRDYKPPIDGVY 168
 DB 125 IEINIOKKKKQSMYIKRLSKNIQTVPILTSEKRLILHAFLVILKKSSILKSL-----L 178
 QY 169 IDEQKGGDKQLTLITCDNYNEKTGVWEKRRKIFVATEVK 206
 DB 179 LMQOYOKTAEVAATSDTYKEKMNDRNSSLFSPFSK 216

RESULT 15
 US-08-137-175A-5
 Sequence 5, Application US/08137175A
 Patent No. 5777095
 GENERAL INFORMATION:
 APPLICANT: BARBOUR, Alan G.
 APPLICANT: BERGSTROM, Sven
 APPLICANT: HANSSON, Lennart
 TITLE OF INVENTION: IMPROVEMENT IN BORRELIA BURGDOFFERI AND
 TITLE OF INVENTION: PROPHYLAXIS
 NUMBER OF SEQUENCES: 22
 CORRESPONDENCE ADDRESS:
 STREET: BROWDY AND NEIMARK
 STREET: 419 Seventh Street, N.W., Suite 300
 CITY: Washington
 STATE: D.C.
 COUNTRY: USA
 ZIP: 20004
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: PatentIn Release #1.0, Version #1.25
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/137,175A
 FILING DATE: 26-OCT-1993
 CLASSIFICATION: 424
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: PCT/US92/08972
 FILING DATE: 22-OCT-1992
 ATTORNEY/AGENT INFORMATION:
 NAME: COOPER, Ivet P.
 REGISTRATION NUMBER: 28,005
 REFERENCE/DOCKET NUMBER: BARBOUR-1B
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: 202-628-5197
 TELEFAX: 202-737-3528
 TELEX: 248633

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 2, 2001, 10:20:33 ; Search time 16.27 Seconds
(without alignments)
432.939 Million cell updates/sec

Title: US-09-292-437-3
Perfect score: 1076
Sequence: 1 MKRWNLMTIAGVILIVA.....YNEKGVMEKRKIFVATEVK 206

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 268485 seqs, 34193795 residues

Total number of hits satisfying chosen parameters: 268485

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :
1: /SIDSI/gcgdata/geneseq/geneseqp/AA1980.DAT:*
2: /SIDSI/gcgdata/geneseq/geneseqp/AA1981.DAT:*
3: /SIDSI/gcgdata/geneseq/geneseqp/AA1982.DAT:*
4: /SIDSI/gcgdata/geneseq/geneseqp/AA1983.DAT:*
5: /SIDSI/gcgdata/geneseq/geneseqp/AA1984.DAT:*
6: /SIDSI/gcgdata/geneseq/geneseqp/AA1985.DAT:*
7: /SIDSI/gcgdata/geneseq/geneseqp/AA1986.DAT:*
8: /SIDSI/gcgdata/geneseq/geneseqp/AA1987.DAT:*
9: /SIDSI/gcgdata/geneseq/geneseqp/AA1988.DAT:*
10: /SIDSI/gcgdata/geneseq/geneseqp/AA1989.DAT:*
11: /SIDSI/gcgdata/geneseq/geneseqp/AA1990.DAT:*
12: /SIDSI/gcgdata/geneseq/geneseqp/AA1991.DAT:*
13: /SIDSI/gcgdata/geneseq/geneseqp/AA1992.DAT:*
14: /SIDSI/gcgdata/geneseq/geneseqp/AA1993.DAT:*
15: /SIDSI/gcgdata/geneseq/geneseqp/AA1994.DAT:*
16: /SIDSI/gcgdata/geneseq/geneseqp/AA1995.DAT:*
17: /SIDSI/gcgdata/geneseq/geneseqp/AA1996.DAT:*
18: /SIDSI/gcgdata/geneseq/geneseqp/AA1997.DAT:*
19: /SIDSI/gcgdata/geneseq/geneseqp/AA1998.DAT:*
20: /SIDSI/gcgdata/geneseq/geneseqp/AA1999.DAT:*
21: /SIDSI/gcgdata/geneseq/geneseqp/AA2000.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	160.5	14.9	251	Y00252	Enterococcus faeca
2	158.5	14.7	284	Y00158	Enterococcus faeca
3	154.5	14.4	197	Y00253	Enterococcus faeca
4	146	13.6	236	Y00159	Enterococcus faeca
5	137	12.7	223	W61218	Streptococcus pneu
6	90	8.4	167	W89768	Streptococcus pneu
7	90	8.4	167	Y20763	Human neurofilamen
8	89.5	8.3	411	W68011	Yeast immunophilin
9	89.5	8.3	1617	Y35617	Chlamydia pneumonia
10	89	8.3	359	W98324	H. pylori GPO 141
11	88.5	8.2	289	W55072	Streptococcus pneu
12	88.5	8.2	309	Y30350	37 kDa pneumococca

13	88.5	8.2	309	20	W82496
14	88.5	8.2	310	21	Y81668
15	87	8.1	310	20	Y00215
16	87	8.1	347	13	Y00214
17	87	8.1	493	13	R26944
18	84.5	7.9	293	20	Y00257
19	84.5	7.9	316	20	Y00256
20	84.5	7.8	170	21	Y84722
21	83.5	7.8	680	12	R12540
22	83.5	7.8	2432	21	Y85565
23	83	7.7	1035	20	W90264
24	82.5	7.7	768	21	Y81596
25	82.5	7.7	783	20	W97416
26	82.5	7.7	991	21	Y83171
27	82.5	7.7	991	21	Y70120
28	82	7.6	235	19	W62784
29	82	7.6	235	19	W62787
30	82	7.6	235	19	W62788
31	82	7.6	307	19	P90366
32	82	7.6	307	19	W80703
33	82	7.6	427	20	Y16590
34	82	7.6	558	13	R27363
35	82	7.6	911	20	Y55957
36	82	7.6	1897	21	Y81785
37	82	7.6	1897	21	Y56100
38	81.5	7.5	783	11	R05804
39	81	7.5	289	20	Y38690
40	80.5	7.5	653	20	Y35021
41	80.5	7.5	759	19	W41927
42	80.5	7.5	759	19	W46593
43	80	7.4	173	18	W28351
44	80	7.4	186	19	W77774
45	80	7.4	273	12	R13141

ALIGNMENTS

RESULT 1	ID	Y00252	standard; Protein: 251 AA.
XX	AC	Y00252;	
XX	DT	20-APR-1999	(first entry)
XX	DE	Enterococcus faecalis	protein EFL30.
XX	KW	Enterococcus faecalis; infection; vaccine; immune response; diagnosis; detection; attenuation; antigenic.	
XX	OS	Enterococcus faecalis.	
XX	PN	W09850554-A2.	
XX	PD	12-NOV-1998.	
XX	PF	04-MAY-1998;	98WO-US08959.
XX	PR	14-NOV-1997;	97US-0066009.
XX	PR	06-MAY-1997;	97US-0044031.
XX	PR	16-MAY-1997;	97US-0046655.
XX	PA	(HUMA-) HUMAN GENOME SCI INC.	
XX	PI	Bailey C, Choi GH, Hromockyj A, Kunsch CA;	
XX	PI	WPI: 1999-070095/06.	
XX	DR	N-PSDB; X20242.	
XX	PT	New isolated Enterococcus faecalis polynucleotides - used to develop products for the detection of Enterococcus and for use in vaccines for prevention or attenuation of Enterococcus infection	
XX	PT		

S. pneumoniae 37-k
Streptococcus pneu
Enterococcus faeca
Enterococcus faeca
P.falcitparum LSA g
Enterococcus faeca
Enterococcus faeca
A decorin binding
Rabbit sperm-speci
Human homologue of
Rat NBC protein.
Streptococcus pneu
A membrane bound p
Cell wall protein
Staph. epidermidis
Streptococcal pyro
Mutant streptococ
Partial streptococ
S. pneumoniae prot
A trigger factor (Sequence of a poly
Mouse STE20-relate
Human protein tyro
LAR tyrosine phosph
C-terminal of nati
Neisseria meningit
Chlamydia pneumoni
Homo sapiens telom
Human telomerase p
staphylococcus aur
staphylococcus aur
B.burgdorferi stra

XX Claim 9; Page 241; 301pp; English.
PS
XX The present sequence represents an antigenic polypeptide fragment
CC isolated from *Enterococcus faecalis*. The present invention describes
CC genes, proteins and antigenic polypeptides isolated from *E. faecalis*.
CC The proteins can be used in vaccines for preventing or attenuating an
CC infection caused by a member of the *Enterococcus* genus in an animal.
CC They can also be used for detecting *Enterococcus* antibodies in a sample.
CC The nucleotide sequences can be used for detecting *Enterococcus* nucleic
CC acids. Products from the present invention can also be used for
CC screening compounds to identify agonists and antagonists of *E. faecalis*
CC protein activity.
CC
XX Sequence 197 AA;
SQ

Query Match 14.4%; Score 154.5; DB 20; Length 197;
Best Local Similarity 27.6%; Pred. No. 1.9e-07;
Matches 40; Conservative 30; Mismatches 72; Indels 3; Gaps 3;

OY 62 KPQIKDKSKVAGYIEIPDADIKPEVYGPATPEQNLNRSFAEENESLDDONISAGHT 121
DB 34 kagfenknlpyigaiaipveinlpitkylsnvailt-gagctmkedqvmgkmyalashr 92
OY 122 FIDRENYOFTNLKAARKGSMYFVKVNETRKYKMTSIRDVKEPTDVGVLDEOKKDKLTU 181
DB 93 ledgvsyl-fsplerckkelilyltdlstvylyktsvkekleptrellddvpqgn-mtlc 150
OY 182 ITCDYNEKTGVWEKRRKIFVATEVK 206
DB 151 itcgdlgalttrialvqgtlaatpik 175

RESULT 4
ID Y00159 standard; Protein; 256 AA.
XX
AC Y00159;
XX
DT 20-APR-1999 (first entry)
XX
DE Enterococcus faecalis antigenic polypeptide fragment EF079.
XX
KW Enterococcus faecalis; infection; vaccine; immune response; diagnosis;
XX detection; attenuation; antigenic.
XX
OS Enterococcus faecalis.
XX
PN W09850554-A2.
XX
PD 12-NOV-1998.
XX
PE 04-MAY-1998; 98WO-US08959.
XX
PR 14-NOV-1997; 97US-0066009.
XX 06-MAY-1997; 97US-0044031.
PR 16-MAY-1997; 97US-0046655.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
PI Bailey C, Choi GH, Hromockyj A, Kunsch CA;
XX
DR MPI: 1999-070095/06.
XX N-PSDB: X20149.
XX
PT New isolated *Enterococcus faecalis* polynucleotides - used to develop
PT products for the detection of *Enterococcus* and for use in vaccines
PT for prevention or attenuation of *Enterococcus* infection
XX
PS Claim 9; Page 170; 301pp; English.
XX
CC The present sequence represents an antigenic polypeptide fragment

CC isolated from *Enterococcus faecalis*. The present invention describes
CC genes, proteins and antigenic polypeptides isolated from *E. faecalis*.
CC The proteins can be used in vaccines for preventing or attenuating an
CC infection caused by a member of the *Enterococcus* genus in an animal.
CC They can also be used for detecting *Enterococcus* antibodies in a sample.
CC The nucleotide sequences can be used for detecting *Enterococcus* nucleic
CC acids. Products from the present invention can also be used for
CC screening compounds to identify agonists and antagonists of *E. faecalis*
CC protein activity.
CC
XX Sequence 256 AA;
SQ

Query Match 13.6%; Score 146; DB 20; Length 256;
Best Local Similarity 25.6%; Pred. No. 1.8e-06;
Matches 46; Conservative 37; Mismatches 83; Indels 14; Gaps 5;

OY 19 VAAVLEAKPHIDNYLHDKDKDEKIEQYDKNVKQEA-----SKDKQQAQKQIPKDKS 70
DB 15 liahyaakasgentekmaeigekmekngelakkgsnpgldfsetqkltkpkdkysfes 74
OY 71 KVAGYIEIPDADIKPEVYGPATPEQNLNRSFAEENE---SLDDONISAGHTFIDRN 127
DB 75 htigvltltpkinvtrlpif-dktnalllekyslilegtsyptggtctnchavisgrjlpqak 133
OY 128 YQFTNLKAARKGSMYFVKVNETRKYKMTSIRDVKEPTDVGVLDEOKKDKLTUICDDY 187
DB 134 l-ftldlpelkkdfeiyngkrlayvqdlkveptdtkdlhiesgd-lyvllictcy 191

RESULT 5
ID W61218 standard; Protein; 223 AA.
XX
AC W61218;
XX
DT 02-OCT-1998 (first entry)
XX
DE Streptococcus pneumoniae SP0093 protein.
XX
KW Streptococcus pneumoniae; antigen; vaccine; infection; diagnosis;
XX detection; pneumonia; otitis media; meningitis.
XX
OS Streptococcus pneumoniae.
XX
FH Key Location/Qualifiers
FT Misc-difference 33
FT /label= unknown
FT /note= "encoded by GNF"
XX
PN W09818930-A2.
XX
PD 07-MAY-1998.
XX
PE 30-OCT-1997; 97WO-US19422.
XX
PR 31-OCT-1996; 96US-0029960.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
PI Choi GH, Hromockyj A, Johnson LS, Kunsch CA;
XX
DR MPI: 1998-272224/24.
XX N-PSDB: V27404.
XX
PT Nucleic acid encoding antigenic peptide(s) from *Streptococcus*
PT pneumoniae - or their epitope-containing fragments, useful in
PT protective or therapeutic vaccines, and for diagnosis
XX
PS Claim 11; Page 82; 118pp; English.
XX
CC The present sequence represents a protein from *Streptococcus pneumoniae*.
CC The nucleic acid sequence encoding the *Streptococcus pneumoniae* protein

CC can be useful in vaccines for inducing protective antibodies against
 CC Streptococcus pneumoniae, for treatment or prevention of infection e.g.
 CC pneumonia, otitis media or meningitis. Probes based on the nucleic acid
 CC are used to detect Streptococcus infection (by usual hybridisation or
 CC amplification methods), also for isolating Streptococcus genes or their
 CC allelic variants. The protein can be used similarly to detect specific
 CC antibodies in standard immunoassays, especially for diagnosing or
 CC monitoring infections. Antibodies which bind the protein are used to
 CC detect corresponding antigens, to purify the protein and for passive
 CC immunisation (optionally coupled to a toxin). Vaccines are administered,
 CC e.g. by injection, orally or through the skin, typically at 0.01-1000
 CC (especially 10-300) mu g/ml per dose.

XX Sequence 223 AA:

Query Match 12.7% Score 137; DB 19; Length 223;
 Best Local Similarity 30.8%; Pred. No. 1,1e-05;
 Matches 45; Conservative 23; Mismatches 62; Indels 16; Gaps 7;

QY 64 QIPRDKSKVAGYIEIPDADIREPVYGPATPEOLNRGVSAEENE-SLDDONIS--IAGH 120
 Db 52 qvsddpavgyislpelmeptylg-adyhlgmgjahvdgfpplpldgqgirsylagh 110
 QY 121 TFIIDRPNVOTNLKAAGKSGMYEFKVGNETRKKYMTSIRBYKPPDVGVLDEQKGRKQOLT 180
 Db 111 r-aepsnvftrhdqjlkvgdalydngelveygmmdteillpsewekleavssakn-lmt 168
 QY 181 LITCD--DYN-----EKTGYWEK 196
 Db 169 litcdpiftrkrllyvferavayqk 194

RESULT 6
 W89768 standard; Protein; 167 AA.

XX 16-MAR-1999 (first entry)
 XX Staphylococcus aureus protein SEQ ID #5216.
 KW Computer readable medium; vaccine; S.aureus infection; immunodetection;
 KW cellulitis; eyelid infection; food poisoning; osteomyelitis; therapy;
 KW skin infection; surgical wound infection; scalded skin syndrome;
 KW toxic shock syndrome.
 XX Staphylococcus aureus.

XX Key Location/Qualifiers
 XX Misc-difference 1..16 /note="these residues represent a line of missing text
 XX in the sequence listing in the specification.
 XX They are included to maintain the residue
 XX numbering given in the specification for this
 XX protein sequence"

XX EP786519-A2.
 XX 30-JUL-1997.
 XX 07-JAN-1997: 97EP-0100117.
 XX 05-JAN-1996: 96US-0009861.
 XX (HUMA-) HUMAN GENOME SCI INC.
 XX Barash SC, Choi GH, Dillon PJ, Fannon MR, Kunsch CA;
 XX Rosen CA;
 XX WPI: 1997-374922/35.
 XX

PT Polynucleotide(s) and proteins derived from Staphylococcus aureus
 PT stored on computer readable medium and used in the production of
 PT anti-S.aureus vaccines

PS Claim 23: Page 3213-3214; 3271pp; English.

XX This sequence represents a Staphylococcus aureus protein sequence of the
 CC invention. The DNA sequences encoding the S.aureus proteins are recorded
 CC on a computer readable medium, preferably selected from a floppy or hard
 CC disk, random access memory (RAM), read-only memory (ROM) or CD-ROM.
 CC Homology searches using the S.aureus DNA sequences allows putative
 CC functions to be assigned so that protein-encoding or regulatory regions
 CC of commercial, therapeutic or industrial importance can be obtained.
 CC Specifically, sequences which are likely to encode antigens have been
 CC identified and these polypeptides can be used in a vaccine composition
 CC against S.aureus infection. The polypeptides can also be used in a kit
 CC for the immunodetection of S.aureus in a sample. S.aureus is implicated
 CC in numerous human diseases, including cellulitis, eyelid infections, food
 CC poisoning, osteomyelitis, skin and surgical wound infections, scalded
 CC skin syndrome, toxic shock syndrome, etc. Organisms transformed with the
 CC DNA sequences can be used for recombinant production of the polypeptides.
 CC The new DNA sequences (and their fragments) are useful as primers or
 CC probes for isolating homologues of any of the 5191 S.aureus DNA sequences
 CC contained on the computer readable medium.

XX Sequence 167 AA:

Query Match 8.4% Score 90; DB 18; Length 167;
 Best Local Similarity 20.8%; Pred. No. 0.26;
 Matches 40; Conservative 30; Mismatches 58; Indels 64; Gaps 4;

QY 5 TNRMTIAGVLIIVAAVYLFAPKPHIDNYLHDKD---DEKIEQYDKNKEQASKD--KKQ 59
 Db 28 takvyslatvllllyglvafalfayvdhskkemlnqkqgqkxkqknaekerrkkq 87
 QY 60 QAKPOIPRDKSKVAGYIEIPDADIREPVYGPATPEOLNRGVSAEENESLDDONISIAG 119
 Db 88 qeekqneldsqangyqqlpqngnygyvppqgqaprk----- 124
 QY 120 HTFIIDRPNVOTNLKAAGKSGMYEFKVGNETRKKYMTSIRBYKPPDVGVLDEQKGRKQOLT 179
 Db 125 ----qrpakeenddkask-----deskdodka 148
 QY 180 TLITCDYNEKT 191
 Db 149 sqdksddnqkkt 160

RESULT 7
 Y20763
 ID Y20763 standard; Protein; 554 AA.

XX Y20763;
 XX 22-JUL-1999 (first entry)
 XX Human neurofilament-M mutant protein fragment 45.

XX Human; beta-amyloid precursor protein; beta-APP; diagnosis; cancer;
 KW frameshift mutation; age-related disease; neurodegenerative disorder;
 KW Alzheimer's disease; Down's syndrome; myotonic dystrophy; neuronal;
 KW Huntington's disease; multiple sclerosis; alcoholic liver disease;
 KW diabetes mellitus type II; microtubule associated protein; Tau; Big Tau;
 KW ubiquitin B; apolipoprotein E; MAP2; neurofilament-L; neurofilament-M;
 KW neurofilament-F; presenilin I; presenilin II; cellular tumour antigen;
 KW glial fibrillary acidic protein; GFAP; p53; semaphorin III; HUPF-1;
 KW bcl-2; B-cell leukemia/lymphoma 2 proto-oncogene; HMGP-C; NSP-A;
 KW high mobility group protein-C; neuroendocrine specific protein A.

XX Synthetic.
 OS Homo sapiens.
 XX

XX 20-NOV-1998; 98WO-IB01890.
 PF 04-NOV-1998; 98US-0107078.
 PR 21-NOV-1997; 97FR-0014673.
 XX (GEST) GENSET.
 PA Grifals R;
 PI MPI: 1999-357842/30.
 DR
 XX Genome sequence of Chlamydia pneumoniae
 PT
 PS Page 1344-1347; Disclosure: 1912pp; English.
 CC Y3584-Y35879 represent the proteins encoded by all the open reading
 CC frames in the complete genome (see X91990) of Chlamydia pneumoniae.
 CC C. pneumoniae causes respiratory disease such as pneumonia and
 CC bronchitis and is thought to be a contributing factor in heart
 CC disease, sarcoidosis, sinusitis, purulent otitis media, erythema
 CC nodosum or pharyngitis. The polypeptides encoded by the open reading
 CC frames of the C. pneumoniae genome (see Y3584-Y35879) can be used in
 CC immunogenic compositions as vaccines. Vectors containing C. pneumoniae
 CC nucleotide sequences can also be used as immunogenic compositions,
 CC especially where the vector directs the expression of a neutralising
 CC epitope of C. pneumoniae.
 CC
 XX Sequence 1617 AA;
 SQ

Query Match 8.3%; Score 89.5; DB 20; Length 1617;
 Best Local Similarity 24.7%; Pred. NO. 7.2;
 Matches 39; Conservative 22; Mismatches 58; Indels 39; Gaps 5;
 QY 49 VKEQSKKKQKQAKPOIKDKSKVAGYIEIPDADIKPEVYEPAPTEQLN----- 98
 DB 153 vsekssdtkerkdletedepksksglkevssdlpkspectavaaisedleisenisardp 212
 QY 99 -RGVSFAEEN--ESLIDONISINAGHTFID--RENYOFTNKAARKGSMVFKYGNETR 151
 DB 213 lqglafyknssqsgisakdsfgilifsgsgangslgfenkakpsgaavy----- 264
 QY 152 KYKMTSINDVKPTDYGVLDEQKGRKQKLTLTCDPNE 189
 DB 265 -----sdrdivfenl-----vkglisfisceel 288

RESULT 10
 W98324
 ID W98324 standard; Protein; 359 AA.
 XX
 AC W98324;
 XX
 DT 31-MAR-1999 (first entry)
 XX
 DE H. pylori GHP0 1411 protein.
 XX
 KW GHP0 protein; Helicobacter infection; gastroduodenal disease; gastritis;
 KW peptic ulcer disease.
 XX
 OS Helicobacter pylori.
 XX
 PN W09843478-A1.
 PD
 PD 08-OCT-1998.
 XX
 PF 01-APR-1998; 98WO-US06371.
 XX
 PR 29-JUL-1997; 97US-0902615.
 PR 01-APR-1997; 97US-0833457.
 PR 24-JUN-1997; 97US-0881227.
 XX

PA (HUMA-) HUMAN GENOME SCI INC.
 PA (INMR) MERIEUX ORAVAX PASTEUR MERIEUX SERUMS.
 XX
 PI Al-Garawi A, Kleanthouse H, Miller C, Oomen RP, Tomb J;
 XX MPI: 1998-542293/46.
 DR
 DR N-PSDB; X14043.
 XX
 PT New isolated Helicobacter polynucleotides - used to develop products
 PT for the diagnosis, prevention and treatment of Helicobacter
 PT infections and gastrointestinal diseases
 XX
 PS Claim 8; Page 435-437; 2054pp; English.
 CC
 CC This sequence represents a Helicobacter pylori GHP0 protein of the
 CC invention. The polypeptides can be used for preventing or treating
 CC Helicobacter infections, and gastroduodenal diseases associated with
 CC these infections, including acute, chronic, and atrophic gastritis, and
 CC peptic ulcer diseases, e.g. gastric and duodenal ulcers. They can also be
 CC used for the production of antibodies. The products can also be used for
 CC detection and diagnosis.
 CC
 XX Sequence 359 AA;
 SQ

Query Match 8.3%; Score 89; DB 19; Length 359;
 Best Local Similarity 25.1%; Pred. NO. 0.97;
 Matches 48; Conservative 20; Mismatches 63; Indels 60; Gaps 10;
 QY 24 FAKPHIDNYLHD--KDKDEKIDQYDKNKVEQSKKKQKQAKPOIKDKSKVAGYIEIP-- 79
 DB 175 ylikhpsgychidlkkaevdvdtkytkpsdevilkrtikltiyevtnk--hckpfe 231
 QY 80 -----DADIKPEVYPG-----PATPEOLNRGVSFAENESLIDONIS 116
 DB 232 neyrnkeegiyvdlttgcpilssadkydsgcgwpsfskypinkdvkydedeslnrkie 291
 QY 117 I-----AGHTFTIDRPN-----YOFTNKAARK--GSMV-YFKYGNETR 151
 DB 292 vlsrighahlgvfnfdgpkelglglyclsaaalrfplkdkmekegygetfpykkq-elk 350
 QY 152 KYKMTSINDVK 162
 DB 351 ky-----lnckk 357

RESULT 11
 W55072
 ID W55072 standard; Protein; 289 AA.
 XX
 AC W55072;
 XX
 DT 02-OCT-1998 (first entry)
 XX
 DE Streptococcus pneumoniae SP0013 protein.
 XX
 KW Streptococcus pneumoniae; antigen; vaccine; infection; diagnosis;
 KW detection; pneumonia; otitis media; meningitis.
 XX
 OS Streptococcus pneumoniae.
 XX
 PN W09818930-A2.
 PD
 PD 07-MAY-1998.
 XX
 PF 30-OCT-1997; 97WO-US19422.
 XX
 PR 31-OCT-1996; 96US-0029960.
 XX
 PA (HUMA-) HUMAN GENOME SCI INC.
 XX
 PI Choi GH, Hromockyj A, Johnson LS, Kunsch CA;
 XX

DR WPI: 1998-272224/24.
 DR N-PSDB: V27333.
 XX Nucleic acid encoding antigenic peptide(s) from Streptococcus
 PT pneumoniae - or their epitope-containing fragments, useful in
 PT protective or therapeutic vaccines, and for diagnosis
 XX
 PS Claim 11; Page 52; 118pp; English.
 CC The present sequence represents a protein from Streptococcus pneumoniae.
 CC The nucleic acid sequence encoding the streptococcus pneumoniae protein
 CC can be useful in vaccines for inducing protective antibodies against
 CC Streptococcus pneumoniae, for treatment or prevention of infection e.g.
 CC pneumonia, otitis media or meningitis. Probes based on the nucleic acid
 CC are used to detect Streptococcus infection (by usual hybridisation or
 CC amplification methods), also for isolating Streptococcus genes or their
 CC allelic variants. The protein can be used similarly to detect specific
 CC antibodies in standard immunoassays, especially for diagnosing or
 CC monitoring infections. Antibodies which bind the protein are used to
 CC detect corresponding antigens, to purify the protein and for passive
 CC immunisation (optionally coupled to a toxin). Vaccines are administered,
 CC e.g. by injection, orally or through the skin, typically at 0.01-1000
 CC (especially 10-300) mu g/ml per dose.
 CC
 SQ Sequence 289 AA;
 Query Match 8.2%; Score 88.5; DB 19; Length 289;
 Best Local Similarity 28.0%; Pred. No. 0.8; Mismatches 53; Indels 55; Gaps 11;
 Matches 49; Conservative 18;
 QY 23 LEAKPHIDNVLHDKDKDEKIEQYDKNVKEQASK-----DKKQAK-POIPKDK----- 69
 DB 129 ifak-nlaxqlsakdpnk-efyeknlkeytdkldkdkeskdkfnkipekklivtseg 186
 QY 70 -----SKVAGYIEIPADIKPEYVPGPAPPEQNLRGVS-----FAENESLDDQ- 113
 DB 187 afkyfskayg---vpsaylweinteegtpedqiklveklrqlkypslive--ssvddrp 241
 QY 114 -----NISAGHTEFIDRPNTQFTNLKAKKGSWYFVGNETRRKYMSTIRD 160
 DB 242 mktvsgdtnlpyaqiftd-----siaegkgdsdy-----smmkynldkiaae 285
 RESULT 12
 Y30350
 ID Y30350 standard; Protein: 309 AA.
 XX Y30350;
 AC Y30350;
 XX
 DT 09-NOV-1999 (first entry)
 XX
 DE 37 kDa pneumococcal surface adhesion A protein (PsaA).
 XX
 KW Pneumococcal surface adhesion A protein; PsaA: monoclonal antibody;
 KW vaccine; Streptococcus pneumoniae infection.
 XX
 OS Streptococcus pneumoniae.
 XX
 PN WO9945121-A1.
 XX
 PD 10-SEP-1999.
 XX
 PF 26-FEB-1999; 99WO-US04326.
 XX
 PR 02-MAR-1998; 98US-0076565.
 XX
 PA (USSH) US DEPT HEALTH & HUMAN SERVICES.
 XX
 PI Ades EW, Carlone GM, Sampson JS, Tharpe JA, Westerink MAJ;
 PI Zeiler JL;
 XX
 DR WPI: 1999-540849/45.

DR N-PSDB: Z10411.
 XX
 PT New peptides corresponding to Streptococcus pneumoniae PsaA, used
 PT for treating or preventing Streptococcus pneumoniae infection in a
 PT subject
 XX
 PS Example 8; Page 53-54; 58pp; English.
 CC The present sequence represents a pneumococcal surface adhesion A protein
 CC (PsaA). The specification describes monoclonal antibodies which bind
 CC epitopes of the PsaA protein (see Y30351-54). These peptides can be
 CC used in vaccines to prevent Streptococcus pneumoniae infections. The
 CC antibodies of the invention can also be used to detect S. pneumoniae in
 CC a sample or individual.
 CC
 SQ Sequence 309 AA;
 Query Match 8.2%; Score 88.5; DB 20; Length 309;
 Best Local Similarity 28.0%; Pred. No. 0.88; Mismatches 53; Indels 55; Gaps 11;
 Matches 49; Conservative 18;
 QY 23 LEAKPHIDNVLHDKDKDEKIEQYDKNVKEQASK-----DKKQAK-POIPKDK----- 69
 DB 149 ifak-nlaxqlsakdpnk-efyeknlkeytdkldkdkeskdkfnkipekklivtseg 206
 QY 70 -----SKVAGYIEIPADIKPEYVPGPAPPEQNLRGVS-----FAENESLDDQ- 113
 DB 207 afkyfskayg---vpsaylweinteegtpedqiklveklrqlkypslive--ssvddrp 261
 QY 114 -----NISAGHTEFIDRPNTQFTNLKAKKGSWYFVGNETRRKYMSTIRD 160
 DB 262 mktvsgdtnlpyaqiftd-----siaegkgdsdy-----smmkynldkiaae 305
 RESULT 13
 W82496
 ID W82496 standard; Protein: 309 AA.
 XX W82496;
 AC W82496;
 DT 04-MAR-1999 (first entry)
 XX
 DE S. pneumoniae 37-kDa surface adhesion A protein.
 XX
 KW Surface adhesion A protein; vaccine; detection; serotype; antibody;
 KW diagnostic; immunoassay; treatment; infection; anti-idiotypic.
 XX
 OS Streptococcus pneumoniae.
 XX
 PN US5854416-A.
 XX
 PD 29-DEC-1998.
 XX
 PF 17-SEP-1996; 96US-0715131.
 XX
 PR 17-SEP-1996; 96US-0715131.
 PR 17-SEP-1991; 91US-0791377.
 PR 04-APR-1994; 94US-0222179.
 XX
 PA (USSH) US DEPT HEALTH & HUMAN SERVICES.
 XX
 PI Ades EW, Carlone GM, Russell H, Sampson JS, Tharpe JA;
 PI N-PSDB: V73914.
 DR WPI: 1999-095007/08.
 XX
 PT Nucleic acid encoding the 37 kDa. surface adhesion A of Streptococcus
 PT pneumoniae - useful diagnostically and for production of
 PT recombinant polypeptides
 XX
 PS Claim 1; Column 33-34; 20pp; English.
 XX

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OW nucleic - nucleic search, using sw model

Run on: March 14, 2001, 06:48:04 ; Search time 62.96 Seconds
(without alignments)
1589.589 Million cell updates/sec

Title: US-09-292-437-2

Perfect score: 621
Sequence: 1 atgaaaaaatgacaatcg.....tagctacagaagtaataaa 621

Scoring table: IDENTITY NUC
Gapop 10.0 , Gapext 1.0

Searched: 280836 seqs, 80580151 residues

Total number of hits satisfying chosen parameters: 561672

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :
1: /cgn2_6/ptodata/1/lna/5A.COMB.seq:*
2: /cgn2_6/ptodata/1/lna/5B.COMB.seq:*
3: /cgn2_6/ptodata/1/lna/6.COMB.seq:*
4: /cgn2_6/ptodata/1/lna/PCrUS.COMB.seq:*
5: /cgn2_6/ptodata/1/lna/backfiles1.seq:*

Pred. NO. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
C 1	65.4	10.5	7218	1	Sequence 14, Appl 5231168
2	59.8	9.5	3095	5	Sequence 1, Appl 5231168-1
3	50.6	8.1	9636	1	Sequence 1, Appl 5231168-1
4	48.4	7.8	4248	3	Sequence 1, Appl 5231168-1
5	48.2	7.8	1368	1	Sequence 5, Appl 5231168-1
6	48.2	7.8	1368	1	Sequence 5, Appl 5231168-1
7	48.2	7.8	1368	1	Sequence 5, Appl 5231168-1
8	48.2	7.8	1425	1	Sequence 5, Appl 5231168-1
9	48.2	7.8	1425	1	Sequence 5, Appl 5231168-1
10	48.2	7.8	1425	1	Sequence 5, Appl 5231168-1
11	48.2	7.8	1524	4	Sequence 7, Appl 5231168-1
12	48.2	7.8	1524	4	Sequence 7, Appl 5231168-1
13	48.2	7.8	1524	4	Sequence 7, Appl 5231168-1
14	48.2	7.8	1524	4	Sequence 7, Appl 5231168-1
15	48.2	7.8	1524	4	Sequence 7, Appl 5231168-1
16	48.2	7.8	1524	4	Sequence 7, Appl 5231168-1
17	46.2	7.4	19124	2	Sequence 13, Appl 5231168-1
18	46.2	7.4	19124	2	Sequence 13, Appl 5231168-1
19	43	6.9	1360	3	Sequence 37, Appl 5231168-1
20	42.6	6.9	1360	3	Sequence 37, Appl 5231168-1
21	41.6	6.7	2042	2	Sequence 11, Appl 5231168-1
22	41.6	6.7	2042	2	Sequence 11, Appl 5231168-1
23	41.6	6.7	2042	2	Sequence 11, Appl 5231168-1
24	41.6	6.7	2042	2	Sequence 11, Appl 5231168-1
25	41.2	6.6	3892	3	Sequence 16, Appl 5231168-1
26	41.2	6.6	3892	3	Sequence 16, Appl 5231168-1
27	41	6.6	1696	3	Sequence 1, Appl 5231168-1
28	40.8	6.6	3022	3	Sequence 217, App 5231168-1
					Sequence 215, App 5231168-1
					Sequence 10, Appl 5231168-1

29	40.4	6.5	3498	3	US-08-293-728-1	Sequence 1, Appl 1
30	40.2	6.5	387	4	PCT-US96-05320A-1342	Sequence 1342, Ap 1
C 31	40.2	6.5	845	2	US-08-743-637B-25	Sequence 25, Appl 1
C 32	40.2	6.5	845	3	US-08-526-840B-25	Sequence 25, Appl 1
33	39.8	6.4	1098	3	US-09-248-335-35	Sequence 35, Appl 1
34	39.8	6.4	3666	2	US-08-682-517-13	Sequence 13, Appl 1
35	39.8	6.4	3666	2	US-08-682-517-13	Sequence 13, Appl 1
36	39.8	6.4	3666	2	US-08-682-517-13	Sequence 13, Appl 1
37	39.8	6.4	4197	2	US-08-682-517-13	Sequence 13, Appl 1
38	39.6	6.4	4197	2	US-08-682-517-13	Sequence 13, Appl 1
39	39.6	6.4	1733	3	US-09-073-569-1	Sequence 1, Appl 1
40	39.6	6.4	4090	3	US-08-569-214-4	Sequence 4, Appl 1
41	38.4	6.2	2007	2	US-08-743-637B-169	Sequence 169, App 1
42	38.4	6.2	2007	2	US-08-526-840B-169	Sequence 169, App 1
43	38.2	6.2	6243	2	US-09-056-075-1	Sequence 1, Appl 1
C 44	38.2	6.2	8457	1	US-07-991-867B-1	Sequence 1, Appl 1
C 45	38.2	6.2	8457	2	US-08-544-332-1	Sequence 1, Appl 1

ALIGNMENTS

RESULT 1
US-08-232-463-14/C
Sequence 14, Application US/08232463
Patent No. 5670367
GENERAL INFORMATION:
APPLICANT: DORNER, F.
APPLICANT: SCHEFFLINGER, F.
APPLICANT: FALKNER, F. G.
TITLE OF INVENTION: RECOMBINANT FOWLPOX VIRUS
NUMBER OF SEQUENCES: 52
CORRESPONDENCE ADDRESS:
ADDRESSEE: Foley & Lardner
STREET: 1800 Diagonal Road, Suite 500
CITY: Alexandria
STATE: VA
COUNTRY: USA
ZIP: 22313-0299
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/232,463
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/07/935,313
FILING DATE:
APPLICATION NUMBER: EP 91 114 300.6
ATTORNEY/AGENT INFORMATION:
NAME: BENT, Stephen A.
REGISTRATION NUMBER: 29,768
REFERENCE/DOCKET NUMBER: 30472/114 IMMU
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703)836-9300
TELEFAX: (703)683-4109
TELEX: 899149
INFORMATION FOR SEQ ID NO: 14:
SEQUENCE CHARACTERISTICS:
LENGTH: 7218 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
IMMEDIATE SOURCE:
CLONE: PTZgpt-Fls
US-08-232-463-14

Query Match 10.5%; Score 65.4; DB 1; Length 7218;

Best Local Similarity 5.3%; Pred. No. 3.9e-07;
Matches 21; Conservative 225; Mismatches 151; Indels 0; Gaps 0

[illegible]

```

RESULT      2
5231168-1
; Patent No. 5231168
; APPLICANT: DZIGIEL, MORTEN;BORRE, MARTIN;JEPSEN, SOREN
; VUDST, JENS;RIENECK, KLAUS;WIND, ANNETTE;JAKOBSEN, PALLE H.
; TITLE OF INVENTION: MALARIA ANTIGEN
; NUMBER OF SEQUENCES: 19
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/409,658
; FILING DATE: 18-SEP-1989
; SEQ ID NO:1:
; LENGTH: 3095
; 5231168-1

```

Query Match	9.6%;	Score 59.8;	DB 5;	Length 3095;
Best Local Similarity	44.4%;	Pred. No. 6.8e-06;		
Matches 241;	Conservative	0;	Mismatches 302;	Indels 0;
			Gaps	0;

[illegible]

Db 1560 agtttgagaatctctcccgagatataaaatgaaaaaggttcaactlgaatlagtagagt 16119

QY 432 taaagtttggfcaatgaaacacgctagatataaatgacaagtataagatgtaagctac 491

Db 1620 tgaagaatctctaccagaagaagataaaaaatgaaaaaggttcaactlgaatgttagagt 16779

QY 492 agatgttagagttctagatgaacaaaaagtaagaataacacattaacatttaacttg 551

Db 1680 tgaagaatctctaccagaagataaaaaatgaaaaaggttcaactlgaatgtagtgagttga 17399

QY 552 tgalgattacaatgaaaaagacagggctgttgggaaaaagctaaatctttagtactaga 611

Db 1740 agaatactctaccagaagataaaaaatgaaaaaggttcaactlgaatgtagtgagttgaaga 17999

QY 612 agt 614

Db 1800 aat 1802

```

RESULT 3
US-08-323-170B-1
Sequence 1: Application US/08323170B
Patent No. 5733772
GENERAL INFORMATION:
APPLICANT: Williamson, Kim C.
APPLICANT: Kaslow, David C.
TITLE OF INVENTION: Cloning and Expression of Plasmodium
TITLE OF INVENTION: falciparum Transmission-Blocking Target
NUMBER OF SEQUENCES: 4
CORRESPONDENCE ADDRESSES:
ADDRESSEE: Townsend and Townsend and Crew LLP
STREET: Two Embarcadero Center, 8th Floor
CITY: San Francisco
STATE: California
COUNTRY: USA
ZIP: 94111-3834
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/323,170B
FILING DATE: 13-OCT-1994
CLASSIFICATION: 424
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/010,409
FILING DATE: 29-JAN-1993
ATTORNEY/AGENT INFORMATION:
NAME: Quine, Jonathan A.
REGISTRATION NUMBER: P-41,261
REFERENCE/DOCKET NUMBER: 015280-113100US
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 576-0200
TELEFAX: (415) 576-0300
INFORMATION FOR SEQ. ID NO. 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 9636 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
FEATURE:
NAME/KEY: CDS
LOCATION: 149..9556
US-08-323-170B-1

```

Query Match	8.18;	Score 50.6;	DB 1;	Length 9636;
Best Local Similarity	43.5%;	Pred. No. 0.0014;		
Matches 230;	Conservative	0;	Mismatches 299;	Indels 0;
Gaps				
Oy	84	tatgcataattatcttcacgataagaagatgcgaagaatctgcacacataatgataaaaa	143	


```

COUNTRY: USA
ZIP: 30303
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentln Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/021,601
FILING DATE: 19930212
CLASSIFICATION: 514
ATTORNEY/AGENT INFORMATION:
NAME: Spratt, Gwendolyn D.
REGISTRATION NUMBER: 36,016
REFERENCE/DOCKET NUMBER: 1414,057
TELECOMMUNICATION INFORMATION:
TELEPHONE: 404/688-0770
TELEFAX: 404/688-9880
INFORMATION FOR SEQ ID NO: 5:
SEQUENCE CHARACTERISTICS:
LENGTH: 1368 base pairs
TYPE: NUCLEIC ACID
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
FEATURE:
NAME/KEY: CDS
LOCATION: 1..1368
US-08-021-601-5

```

```

Query Match          7.8%  Score 48.2; DB 1; Length 1368;
Best Local Similarity 44.9%  Pred. No. 0.0032;
Matches 182; Conservative 0; Mismatches 223; Indels 0; Gaps 0;

```

```

QY 106 aagaataaagtgtaagatggaacatataataaataaagaacagcgagtaaa 165
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 121 AAGCATTGTGTAATAAGTAAGTAAGGAGGAGGAGCTGTTAAAGAGGACGAGAA 180
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 166 gataaaagcagacagcttaaacctcaatccgaagaataatcgaagtgcgagctat 225
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 181 AAGCTACTTGAGAAAGTACCTGATGTTTACAGATGTTAAAGCATTTGGAGGAAG 240
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 226 atgaataatccagatgctgataataaagaacagatataccagagccgaacaacctgaa 285
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 241 ATATATATTTGATGATGATGATTAACAAACATATATCTTTAGAACCATTTATCGAAGAT 300
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 286 caattaataagagtgtaagctttgcgaagaataatgaatcagatagatcaataatatt 345
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 301 AAGAAAAAATAAAGACATTTATGGAAGATGCTTTATTACATGAACATTATGATATAT 360
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 346 tcaatgcagagacacactttcaatgcacgctcgaacatacaattacaatcttaagca 405
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 361 GCAGAAAGAGATGTGAACCCGTTCTGTATCAATCTTCGAGAGTTATGTAAGAAAT 420
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 406 gccaaaaaagtgatggtggtacttaagttgtaatagaacacgtaagtaataatg 465
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 421 ACTGAAGAGGACTGAACGTTTATTATGAAGATATGTAAGATATATCAAGGATATTTTA 480
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 466 acaagataagagtgtaagcctcaacagatgtaagttcagat 510
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 481 AGTAAATTAATCAACATATCAAGAAATTTTATGATGATTAAT 525
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

RESULT 6

```

US-08-082-849B-5
Sequence 5, Application US/08082849B
Patent No. 5677274
GENERAL INFORMATION:
APPLICANT: Leppia, Stephen H.
APPLICANT: Klimpel, Kurt R.
APPLICANT: Ariora, Naveen
APPLICANT: Singh, Yogendra

```

```

APPLICANT: Nichols, Peter J.
TITLE OF INVENTION: Anthrax Toxin Fusion Proteins and
TITLE OF INVENTION: Related Methods
NUMBER OF SEQUENCES: 35
CORRESPONDENCE ADDRESS:
ADDRESS: Townsend and Townsend and Crew LLP
STREET: Two Embarcadero Center, Eighth Floor
CITY: San Francisco
STATE: California
COUNTRY: USA
ZIP: 94111-3834
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentln Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/082,849B
FILING DATE: 25-JUN-1993
CLASSIFICATION: 514
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/021,601
FILING DATE: 12-FEB-1993
ATTORNEY/AGENT INFORMATION:
NAME: Weber, Kenneth A.
REGISTRATION NUMBER: 31,677
REFERENCE/DOCKET NUMBER: 15280-161-1
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 576-0200
TELEFAX: (415) 576-0300
INFORMATION FOR SEQ ID NO: 5:
SEQUENCE CHARACTERISTICS:
LENGTH: 1368 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
HYPOTHETICAL: NO
ANTI-SENSE: NO
ORIGINAL SOURCE:
ORGANISM: Bacillus anthracis
FEATURE:
NAME/KEY: CDS
LOCATION: 1..1368
OTHER INFORMATION: /product= "L(F(1-254))-TR-DE(401-602)"
US-08-082-849B-5

```

```

Query Match          7.8%  Score 48.2; DB 1; Length 1368;
Best Local Similarity 44.9%  Pred. No. 0.0032;
Matches 182; Conservative 0; Mismatches 223; Indels 0; Gaps 0;

```

```

QY 106 aagaataaagtgtaagatggaacatataataaataaagaacagcgagtaaa 165
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 121 AAGCATTGTGTAATAAGTAAGTAAGGAGGAGGAGCTGTTAAAGAGGACGAGAA 180
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 166 gataaaagcagacagcttaaacctcaatccgaagaataatcgaagtgcgagctat 225
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 181 AAGCTACTTGAGAAAGTACCTGATGTTTACAGATGTTAAAGCATTTGGAGGAAG 240
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 226 atgaataatccagatgctgataataaagaacagatataccagagccgaacaacctgaa 285
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 241 ATATATATTTGATGATGATGATTAACAAACATATATCTTTAGAACCATTTATCGAAGAT 300
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 286 caattaataagagtgtaagctttgcgaagaataatgaatcagatagatcaataatatt 345
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 301 AAGAAAAAATAAAGACATTTATGGAAGATGCTTTATTACATGAACATTATGATATAT 360
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 346 tcaatgcagagacacactttcaatgcacgctcgaacatacaattacaatcttaagca 405
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 361 GCAGAAAGAGATGTGAACCCGTTCTGTATCAATCTTCGAGAGTTATGTAAGAAAT 420
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 406 gccaaaaaagtgatggtggtacttaagttgtaatagaacacgtaagtaataatg 465
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

Db 421 ACTGAAAAGCGCTGACCTTTATTTATGAAATAGATAGATATTAACCAAGGATATTTTA 480
Qy 466 acaagataagagatgtaagcctacagatgtaagcttcagat 510
Db 481 AGTAAATTAATCAACCATATCAGAAATTTTATGATGTATTAAAT 525

RESULT 7
PCT-US94-01624-5

Sequence 5, Application PC/TUS9401624
GENERAL INFORMATION:
APPLICANT: Leppla, Stephen H.
APPLICANT: Klimpel, Kurt R.
APPLICANT: Aroia, Naveen
APPLICANT: Singh, Yogendra
APPLICANT: Nichols, Peter J.
TITLE OF INVENTION: ANTHRAX TOXIN FUSION PROTEINS AND
NUMBER OF SEQUENCES: 31
TITLE OF INVENTION: RELATED METHODS
CORRESPONDENCE ADDRESS:
ADDRESS: TOWNSEND and TOWNSEND KHOURIE and CREM
STREET: Steuart Street Tower, 20th Floor, One Market
STREET: Plaza
CITY: San Francisco
STATE: CA
COUNTRY: USA
ZIP: 94105
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US94/01624
FILING DATE: June 25, 1993
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: Weber, Kenneth A.
REGISTRATION NUMBER: 31,677
REFERENCE/DOCKET NUMBER: 15280-115
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 543-9600
TELEFAX: (415) 543-5043
INFORMATION FOR SEQ ID NO: 5:
SEQUENCE CHARACTERISTICS:
LENGTH: 1368 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
HYPOTHETICAL: NO
ANTI-SENSE: NO
ORIGINAL SOURCE:
ORGANISM: Bacillus anthracis
FEATURE:
NAME/KEY: CDS
LOCATION: 1..1368
OTHER INFORMATION: /Product=
OTHER INFORMATION: "Lf(1-254)--TR--PE(401-602)"
PCT-US94-01624-5

Query Match 7.8%; Score 48.2; DB 4; Length 1368;
Best Local Similarity 44.9%; Pred. No. 0.0032;
Matches 182; Conservative 0; Mismatches 223; Indels 0; Gaps 0;

Qy 106 aaagataaagatgaaagatgtaaacatagtataaagatgtaaacagagcgagtaaa 165
Db 121 AAACACATGTGTAATAATAGATAAAAGGAGCAAGCTGTAAATAAAAGACGACGACAGAA 180
Qy 166 gataaagaagcagacgtacacctaactcgaataaatacgaagtgccaggtat 225

Db 181 AACCTACTGAGAAAGTACCATCTGATGTTTAGAGATGTATTAACCAATTGGAGGAAG 240
Qy 226 attgaattccagatgctgatatlaaagaaccagatatatccagaccagcaacactgaa 285
Db 241 ATATATATTGTGATGTGATGATATATTAACAAACATATATCTTTAGAACATTATGTGAAGT 300
Qy 286 caattaataagaggtgtaagccttgcaagaagaataatgaatcactagatgatcaaatat 345
Db 301 AAGAAAAAATAATAAACACATTTATGGAAAAAGATCTTTATACATGACATTTATGTATAT 360
Qy 346 tcaatgcygacacacattcatatgaccgtccgaactatcaatcattcaatcctaagca 405
Db 361 GCAAAAGAAAGATATGACCCGTACTTGTATCCAAATCTTCGAGAGATTATGTGAANAAT 420
Qy 406 gccaaaaagtgatggtgtaacttaaaagtgtgaatgaacacgtaagataaagtg 465
Db 421 ACTGAAAAGCGCTGACCTTTATTTATGAAATAGATAGATATTAACCAAGGATATTTTA 480
Qy 466 acaagataagagatgtaagcctacagatgtaagcttcagat 510
Db 481 AGTAAATTAATCAACCATATCAGAAATTTTATGATGTATTAAAT 525

RESULT 8
US-08-021-601-7
Sequence 7, Application US/08021601
Patent No. 5591631
GENERAL INFORMATION:

APPLICANT: Leppla, Stephen H.
APPLICANT: Klimpel, Kurt R.
APPLICANT: Nichols, Peter J.
APPLICANT: Aroia, Naveen
APPLICANT: Singh, Yogendra
TITLE OF INVENTION: ANTHRAX TOXIN FUSION PROTEINS AND
NUMBER OF SEQUENCES: 12
CORRESPONDENCE ADDRESS:
ADDRESSEE: Needle & Rosenberg, P.C.
STREET: 133 Carnegie Way, Suite 400
CITY: Atlanta
STATE: Georgia
COUNTRY: USA
ZIP: 30303
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/021,601
FILING DATE: 19930212
CLASSIFICATION: 514
ATTORNEY/AGENT INFORMATION:
NAME: Spratt, Gwendolyn D.
REGISTRATION NUMBER: 36,016
REFERENCE/DOCKET NUMBER: 1414.057
TELECOMMUNICATION INFORMATION:
TELEPHONE: 404/688-0770
TELEFAX: 404/688-9880
INFORMATION FOR SEQ ID NO: 7:
SEQUENCE CHARACTERISTICS:
LENGTH: 1425 base pairs
TYPE: NUCLEIC ACID
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
FEATURE:
NAME/KEY: CDS
LOCATION: 1..1416
US-08-021-601-7

Query Match 7.8%; Score 48.2; DB 1; Length 1425;

Best Local Similarity 44.9%; Pred. No. 0.0032;
Matches 182; Conservative 0; Mismatches 223; Indels 0; Gaps 0.

QY	106	aaagataaagatctggaagaattgcaacaatcgtatbaaaatctaaagaacagcgagat	1655
Db	130	AAACACATTGTAAAAATATGAAAGTAAAGGGGAGCAACCTGTTAAAAAAGGCGACGAGAA	1691
QY	166	gataaaagcagcaagcttaaacctcaaatccgaaagaataatcgaagtggcaggtat	2225
Db	190	AAGCTACTTGAGAAAGTACCATCTGATGTTTACAGATGTATTAAGCAATTTGGAGCAAG	2439
QY	226	attgaattccagatgcgtgatattaaagaaccgatatccagagcagcaaacctgaa	2655
Db	250	ATATATTTGTGGATGGTGTGATTTACAAAACATATATCTTTAGAACGATTATCTGAAGAT	3091
QY	286	caatlaataagagtgtaagcctttgcgaagaanaatgaatcactagatgtatcaaaatatt	3454
Db	310	AAGAAAAAAATAAAGACATTTATGGAAGAAAGATGCTTTATTACTGTGACATTTATGTATAT	3685
QY	346	tcaattgcagagacacacttccatctgaacgcgtcggaactcatcaattacaacatctlaagca	4059
Db	370	GCAAAAGAGAGATGTGAACCGGTACTTGTATCCAAATCTTGCGAAGATTATGTAGAAAT	4291
QY	406	gccaaaaaagtgtagtgcgtactttaaagcttggttaatgaacaagctlaagtaaaatg	4655
Db	430	ACTGAAAGGCACTGAACGTTTATTATGAATAATGGTAAAGATATTATCAAGGATATTTTA	4889
QY	466	acaagtataagagatgtgtaagccacagatgtaggagttcttagat	510
Db	490	AGTAAATTTATCAACCATATCGAAATTTTGTAGATGTATTAAT	534

RESULT 9
US-08-082-849B-7
Sequence 7, Application US/08082849B
Patent No. 5677274
GENERAL INFORMATION:
APPLICANT: Leppla, Stephen H.
APPLICANT: Klimpel, Kurt R.
APPLICANT: Arora, Navendra
APPLICANT: Singh, Yogendra
APPLICANT: Nichols, Peter J.
TITLE OF INVENTION: Anthrax Toxin Fusion Proteins and
TITLE OF INVENTION: Related Methods
NUMBER OF SEQUENCES: 35
CORRESPONDENCE ADDRESS:
ADDRESSEE: Townsend and Townsend and Crew LLP
STREET: Two Embarcadero Center, Eighth Floor
CITY: San Francisco
STATE: California
COUNTRY: USA
ZIP: 94111-3834
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/082,849B
FILING DATE: 25-JUN-1993
CLASSIFICATION: 514
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/021,601
FILING DATE: 12-FEB-1993
ATTORNEY/AGENT INFORMATION:
NAME: Weber, Kenneth A.
REGISTRATION NUMBER: 31,677
REFERENCE/DOCKET NUMBER: 15280-161-1
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 576-0200
TELEFAX: (415) 576-0300
INFORMATION FOR SEQ ID NO: 7:
SEQUENCE CHARACTERISTICS:

```

? LENGTH: 1425 base pairs
? TYPE: nucleic acid
? STRANDEDNESS: single
? TOPOLOGY: linear
? MOLECULE TYPE: DNA (genomic)
? HYPOTHETICAL: NO
? ORIGINAL SOURCE:
? ORGANISM: Bacillus anthracis
? FEATURE:
? NAME/KEY: CDS
? LOCATION: 1..1416
? OTHER INFORMATION: /product= "LF(1-254)--IR--PE(398-613)"
?
US-08-082-849E-7

```

Query Match 7.8%; Score 48.2; DB 1; Length 1425;
Best Local Similarity 44.9%; Pred. No. 0.0032;
Matches 182; Conservative 0; Mismatches 223; Indels 0; Gaps 0.

QY	106	aagatgaagaatcgaagaattcgacaacatcgataaanaatgtaaagaacagcgagatga	165
Db	130	AAACACATTTGTAAAAATATGAAAGTTAAAGGGGAGCAACCTGTTAAAAAAGGCGCACGAGAA	139
QY	166	gataaaaagcagcaagcgtlaaacctcaaatctcgaaagaataatcgaaaglygcaggtat	225
Db	190	AAGCTACTTGAGAAAGTACCATCTGATGTTTGTAGAGATGATATAAGCAATTTGGAGGAAG	249
QY	226	attgaaattccagatgcttgatataaagaaccgatatatccagagaccagcaaacctgaa	285
Db	250	ATATATATTTGTGATGGTGATATTACAAAACATATACTTTAGAGGATTTATGTGAAGAT	309
QY	286	caatlaaatgaggtgtgaagcttgcagaagaanaatgaacacacagatgaatcaaatatc	345
Db	310	AAGCAAAAAATTAAGACGATTTATGGGAAAGATGCTTTATTACTGTGACACTTATGTATAT	369
QY	346	tcaattgcagagacacacttccatgcacgcgtcggaactatcaatltcaaaacttcaagca	405
Db	370	GCAAAAGAGATATGAAACCCGTACTGTATCCAAATCTTCGGAGATTTATGAGAAAT	429
QY	406	gccaaaaaagtgatgtagtgcgtactttaaagttggtgaatgaacaacgtaagtataaatg	465
Db	430	ACTGMAAAGGACACTGAAACGTTTATTATGAATAATAGTATAGATATTATTCMAAGGATATTTTA	489
QY	466	acaagatlaagagatgttaagcctacagatgaglgaggatcttagat	510
Db	490	AGTAAATTTATCAACCATATCCAGAAATTTTGTAGATCTATTAAT	534

```

RESULT 10
PCT-US94-01624-7
; Sequence 7, Application PC/TUS9401624
; GENERAL INFORMATION:
; APPLICANT: Leppia, Stephen H.
; APPLICANT: Klimpel, Kurt R.
; APPLICANT: Ariora, Naveen
; APPLICANT: Singh, Yogendra
; APPLICANT: Nichols, Peter J.
; TITLE OF INVENTION: ANTHRAX TOXIN FUSION PROTEINS AND
; TITLE OF INVENTION: RELATED METHODS
; NUMBER OF SEQUENCES: 31
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: TOWNSEND AND TOWNSEND KHOURIE and CREW
; STREET: Stewart Street Tower, 20th Floor, One Market
; STREET: Plaza
; CITY: San Francisco
; STATE: CA
; COUNTRY: USA
; ZIP: 94105
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25

```

```

CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US94/01624
FILING DATE: June 25, 1993
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: Weber, Kenneth A.
REGISTRATION NUMBER: 31,677
REFERENCE/DOCKET NUMBER: 15280-115
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 543-9600
TELEFAX: (415) 543-5043
INFORMATION FOR SEQ ID NO: 7:
SEQUENCE CHARACTERISTICS:
LENGTH: 1425 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
HYPOTHEICAL: NO
ORIGINAL SOURCE:
ORGANISM: Bacillus anthracis
FEATURE:
NAME/KEY: CDS
LOCATION: 1..1416
OTHER INFORMATION: /product=
OTHER INFORMATION: "LF(1-254)--TR--PE(398-613)"
PCT-US94-01624-7

```

```

Query Match      7.8%; Score 48.2; DB 4; Length 1425;
Best Local Similarity 44.9%; Pred. No. 0.0032;
Matches 182; Conservative 0; Mismatches 223; Indels 0; Gaps 0;

```

```

QY 106 aagaataaagatgaagaattgaacaatattgaataatgttaagaagaacagcgagtaaa 165
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 130 AACACATTGTAAATGAAGTAAAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 189
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 166 gataaaagcagcaagctcaatccgaatccgaatccgaatccgaatccgaatccgaatcc 225
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 190 AAGCTACTGTGAAGAGTACCATCTGATGTTTGAAGATGTATTAAGCATTTGAGAGAAAG 249
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 226 attgaattccagatgctgataatgaagaacagatataccagagccagcaaacctgaa 285
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 230 ATATATATTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 309
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 286 caataaataagagtgtaagctttgcagaagaataatgaatccatcagatgaatcaaatatt 345
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 310 AAGAAATAATTAAGACATTTATGGAAGATGCTTTTATTAATCAATGACATTTATGTATAT 369
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 346 tcaattgcagacacattcattgacgctccgaactatcaattcaaatcttaagaaga 405
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 370 GCAAAAGAGATATGAACCCGCTACTGTATATCAATCTTCGAGAGATATATAGAAAT 429
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 406 gccaaaaaagtagtattgcttaagaattgtaaatgaagaacagtaagataaataatg 465
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 430 ACTGAAGGCGCTGACCTTTATTTATGAAATAGGTAAGATTTATCAAGCATATTTTTA 489
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 466 acaagataaagatgtaagcctacagatgtaagagatcagat 510
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 490 AGTAAATTAATCAACCATATCAGAAATTTTATGATGATTAAT 534
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

```

RESULT 11
US-08-021-601-9
Sequence 9, Application US/08021601
Patent No. 5591631
GENERAL INFORMATION:
APPLICANT: Leppla, Stephen H.
APPLICANT: Kimpel, Kurt R.
APPLICANT: Nichols, Peter J.
APPLICANT: Arora, Naveen
APPLICANT: Singh, Yogendra
TITLE OF INVENTION: ANTHRAX TOXIN FUSION PROTEINS AND

```

```

TITLE OF INVENTION: RELATED METHODS
NUMBER OF SEQUENCES: 12
CORRESPONDENCE ADDRESS:
ADDRESSEE: Needle & Rosenberg, P.C.
STREET: 133 Carnegie Way, Suite 400
CITY: Atlanta
STATE: Georgia
COUNTRY: USA
ZIP: 30303
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/021,601
FILING DATE: 19930212
CLASSIFICATION: 514
ATTORNEY/AGENT INFORMATION:
NAME: Spratt, Gwendolyn D.
REGISTRATION NUMBER: 36,016
REFERENCE/DOCKET NUMBER: 1414.057
TELECOMMUNICATION INFORMATION:
TELEPHONE: 404/688-0770
TELEFAX: 404/688-9880
INFORMATION FOR SEQ ID NO: 9:
SEQUENCE CHARACTERISTICS:
LENGTH: 1524 base pairs
TYPE: NUCLEIC ACID
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
FEATURE:
NAME/KEY: CDS
LOCATION: 1..1524
US-08-021-601-9

```

```

Query Match      7.8%; Score 48.2; DB 1; Length 1524;
Best Local Similarity 44.9%; Pred. No. 0.0033;
Matches 182; Conservative 0; Mismatches 223; Indels 0; Gaps 0;

```

```

QY 106 aagaataaagatgaagaattgacaatattgaataatgttaagaagaacagcgagtaaa 165
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 121 AACACATTGTAAATTAAGATTAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 180
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 166 gataaaagcagcaagctcaatccgaatccgaatccgaatccgaatccgaatccgaatcc 225
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 181 AAGCTACTGTGAAGAGTACCATCTGATGTTTGAAGATGTATTAAGCATTTGAGAGAAAG 240
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 226 attgaattccagatgctgataatgaagaacagatataccagagccagcaaacctgaa 285
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 241 ATATATATTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 300
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 286 caataaataagagtgtaagctttgcagaagaataatgaatccatcagatgaatcaaatatt 345
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 301 AAGAAATAATTAAGACATTTATGGAAGATGCTTTTATTAATCAATGACATTTATGTATAT 360
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 346 tcaattgcagacacattcattgacgctccgaactatcaattcaaatcttaagaaga 405
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 361 GCAAAAGAGATATGAACCCGCTACTGTATATCAATCTTCGAGAGATATATAGAAAT 420
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 406 gccaaaaaagtagtattgcttaagaattgtaaatgaagaacagtaagataaataatg 465
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 421 ACTGAAGGCGCTGACCTTTATTTATGAAATAGGTAAGATTTATCAAGCATATTTTTA 480
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 466 acaagataaagatgtaagcctacagatgtaagagatcagat 510
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 481 AGTAAATTAATCAACCATATCAGAAATTTTATGATGATTAAT 525
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

```

RESULT 12
US-08-082-849B-9

```

```

: Sequence 9 Application US/08082849B
: Patent No. 5677274
:
: GENERAL INFORMATION:
: APPLICANT: Leppla, Stephen H.
: APPLICANT: Kilmpel, Kurt R.
: APPLICANT: Arora, Naveen
: APPLICANT: Singh, Yogendra
: APPLICANT: Nichols, Peter J.
: TITLE OF INVENTION: Anthrax Toxin Fusion Proteins and
: TITLE OF INVENTION: Related Methods
: NUMBER OF SEQUENCES: 35
:
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Townsend and Townsend and Crew LLP
: STREET: Two Embarcadero Center, Eighth Floor
: CITY: San Francisco
: STATE: California
: COUNTRY: USA
: ZIP: 94111-3834
:
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: PatentIn Release #1.0, Version #1.30
:
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/082.849B
: FILING DATE: 25-JUN-1993
: CLASSIFICATION: 514
:
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 08/021.601
: FILING DATE: 12-FEB-1993
: ATTORNEY/AGENT INFORMATION:
: NAME: Weber, Kenneth A.
: REGISTRATION NUMBER: 31,677
: REFERENCE/DOCKET NUMBER: 15280-161-1
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (415) 576-0200
: TELEFAX: (415) 576-0300
:
: INFORMATION FOR SEQ ID NO: 9:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 1524 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: single
: TOPOLOGY: linear
: MOLECULE TYPE: DNA (genomic)
: HYPOTHEetical: NO
: ORIGINAL SOURCE:
: ORGANISM: Bacillus anthracis
:
: FEATURE:
: NAME/KEY: CDS
: LOCATION: 1..1524
: OTHER INFORMATION: /product= "Lf(1-254)--TR--PE(362-613)"
:
: US-08-082-849B-9
:
: Query Match 7.8%; Score 48.2; DB 1; Length 1524;
: Best Local Similarity 44.9%; Pred. No. 0.0033;
: Matches 182; Conservative 0; Mismatches 223; Indels 0; Gaps 0;
:
: QY 106 aaagataaagatgtaaaagattggaacaatatgataaaaatgtaaaagaacaggcgagtaaa 165
:   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
: Db 121 AAACACATTGTAATAAATGGAAGTAAAGGGGAGGAGCTGTAAAAAAGAGGACACAGAA 180
:
: QY 166 gataaaagcgagcagcctaaaccttaaatccgagaagataaatacgaaaggtgcaggcctat 225
:   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
: Db 181 AAGCTACTTGAGAAAGTACCATCTGATCTTTTAGAGATGTATAAAGCANTTGGAGGAAG 240
:
: QY 226 atcgaattcagatgctgatatataaagaacagtatataccagagcagcaacactgaa 285
:   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
: Db 241 ATATATATTGTGATGTGATATTACAAACATATATCTTTGAGAGCAATTATCTGAAGAT 300
:
: QY 286 caattaaatagagtgtaagctttgcagagaagaataatgatacctctgagatcaataatct 345
:   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
: Db 301 AAGCAAAAATAAAGACACATTATGGAAGAGTGCCTTATTATACATGAACATTATGTATAT 360

```

```

Qy      346   |aa|t|g|g|a|c|a|c|a|c|t|t|a|t|g|a|c|c|g|c|g|a|c|a|t|c|a|t|t|a|c|a|a|c|t|t|a|a|g|c|a|    409
          |||||         |||||         |||||         |||||         |||||         |||||         |||||
Db      361   |G|C|A|A|A|G|A|G|T|A|G|A|C|C|G|T|A|C|T|T|A|T|C|A|A|T|T|T|G|G|A|G|A|T|T|A|T|G|A|G|A|A|T|    420
          |||||         |||||         |||||         |||||         |||||         |||||         |||||
Qy      406   |g|c|c|a|a|a|a|a|g|t|a|t|g|t|g|t|a|c|t|t|a|a|g|t|g|t|a|a|g|t|g|a|a|c|a|c|g|t|a|g|t|a|a|a|t|g|    465
          |||||         |||||         |||||         |||||         |||||         |||||         |||||
Db      421   |A|C|T|G|A|A|A|G|G|C|G|C|G|T|A|C|T|T|A|T|A|T|G|A|A|T|A|G|T|A|G|A|T|A|T|A|T|C|A|A|G|G|A|T|T|T|T|A|    480
          |||||         |||||         |||||         |||||         |||||         |||||         |||||
Qy      466   |a|c|a|a|g|t|a|a|g|a|t|g|t|a|a|g|c|c|t|a|c|a|g|t|a|g|a|g|t|c|t|a|g|t|    510
          |||||         |||||         |||||         |||||         |||||         |||||         |||||
Db      481   |A|G|T|A|A|A|T|T|A|T|C|A|C|A|C|A|T|A|T|C|A|G|A|A|T|T|T|T|A|G|A|T|G|A|T|T|A|A|T|    525
          |||||         |||||         |||||         |||||         |||||         |||||         |||||

RESULT 13
PCT-US94-01624-9
Sequence 9, Application PC/TUS9401624
GENERAL INFORMATION:
APPLICANT: Lepela, Stephen H.
APPLICANT: Klimpel, Kurt R.
APPLICANT: Alora, Naveen
APPLICANT: Singh, Yogendra
APPLICANT: Nichols, Peter J.
TITLE OF INVENTION: ANTHRAX TOXIN FUSION PROTEINS AND
TITLE OF INVENTION: RELATED METHODS
NUMBER OF SEQUENCES: 31
CORRESPONDENCE ADDRESS:
ADDRESSEE: TOWNSEND and TOWNSEND KHOURIE and CREW
STREET: Steuart Street Tower, 20th Floor, One Market
STREET: Plaza
CITY: San Francisco
STATE: CA
COUNTRY: USA
ZIP: 94105
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentln Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US94/01624
FILING DATE: June 25, 1993
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: Weber, Kenneth A.
REGISTRATION NUMBER: 31,677
REFERENCE/DOCKET NUMBER: 15280-115
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 543-9600
TELEFAX: (415) 543-5043
INFORMATION FOR SEQ ID NO: 9:
SEQUENCE CHARACTERISTICS:
LENGTH: 1524 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
HYPOTHETICAL: NO
ORIGINAL SOURCE:
ORGANISM: Bacillus anthracis
FEATURE:
NAME/KEY: CDS
LOCATION: 1..1524
OTHER INFORMATION: /product=
OTHER INFORMATION: "LF(1-254)--R--PE(362-613)"
PCT-US94-01624-9

Query Match              7.8% Score 48.2; DB 4; Length 1524;
Best Local Similarity 44.9%; Pred. No. 0.0033;
Matches 182; Conservative 0; Mismatches 223; Indels 0; Gaps

106 aaagataagaatgtaagaattgaacaatatgatataaaaagtlaaagaacaggcgagttaa 165
```

```

Db 121 AACACATTGTAAATAATGCAACTAAAGGGGAGAGAACTGTTAAAAAGAGCGACGAGAA 180
Oy 166 gataaaagcagcaagcctaaccctaactccgaagaataatcgaaagtgccagctat 225
Db 181 AAGCTACTTGAGAAAGTACCATCTGATGTTTAGACATGTATAAAGCAATTTGGAGAAAG 240
Oy 226 attgaattccagatgctgatatataaagaaccagatataccagaaccgaacacccgtaa 285
Db 241 ATATATATTGTGATGATGATATACAAACATATATCTTTAGAAAGCATTTATCTGAAGAT 300
Oy 286 caattaataagaggtgtaagcttccgagaagaataatgaatcaactagatgatacaaatat 345
Db 301 AAGAAAAAATTAAGACATTTATGCGAAGATGCTTTATTCATGACATGATATGTATAT 360
Oy 346 tcaattgcagagcacacttcaatlgacgctccgaactatacaattacaacttcaaga 405
Db 361 GCAAAAGAAAGGATATGACACCCCTACTGTAATCCATCTTCGGAAGATTATAGAAAAT 420
Oy 406 gccaaaaaagtagtagtggtaacttaaaagtgtgaatgaacacgtaagataaaatg 465
Db 421 ACTGAAAGGCGACTGACGTTTATGAAATAGATATATTCAGAGGATATTTTA 480
Oy 466 acaagataagagatgtaagcctacagatgtaagagttctagat 510
Db 481 AGTAAATTAATCAACCATATCAGAAATTTTATGATTAAT 525

```

RESULT 14 US-08-021-601-1

Sequence 1, Application US/08021601
Patent No. 5591631

GENERAL INFORMATION:
APPLICANT: Leppla, Stephen H.
APPLICANT: Klimpel, Kurt R.
APPLICANT: Nichols, Peter J.
APPLICANT: Atora, Naveen
TITLE OF INVENTION: ANTHRAX TOXIN FUSION PROTEINS AND
NUMBER OF SEQUENCES: 12
CORRESPONDENCE ADDRESS:
ADDRESSEE: Needle & Rosenberg, P.C.
STREET: 133 Carnegie Way, Suite 400
CITY: Atlanta
STATE: Georgia
COUNTRY: USA

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/021.601
FILING DATE: 19930212
CLASSIFICATION: 514
ATTORNEY/AGENT INFORMATION:
NAME: Spratt, Gwendolyn D.
REGISTRATION NUMBER: 36,016
REFERENCE/DOCKET NUMBER: 1414.057
TELECOMMUNICATION INFORMATION:
TELEPHONE: 404/688-0770
TELEFAX: 404/688-9880
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 3291 base pairs
TYPE: NUCLEIC ACID
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
ORGANISM: Bacillus anthracis

FEATURE:
NAME/KEY: CDS
LOCATION: 580..2907
US-08-021-601-1
Query Match
Best Local Similarity 7.8%; Score 48.2; DB 1; Length 3291;
Matches 182; Conservative 0; Mismatches 223; Indels 0; Gaps 0;

```

Oy 106 aaagataaagatgaaagattgaacatatatgataaaatgtaaaagaacagcgagtaaa 165
Db 700 AACACATTGTAAATAATGCAACTAAAGGGGAGAGAACTGTTAAAAAGAGCGACGAGAA 759
Oy 166 gataaaagcagcaagcctaaccctaactccgaagaataatcgaaagtgccagctat 225
Db 760 AAGCTACTTGAGAAAGTACCATCTGATGTTTAGACATGTATAAAGCAATTTGGAGAAAG 819
Oy 226 attgaattccagatgctgatatataaagaaccagatataccagaaccgaacacccgtaa 285
Db 820 ATATATATTGTGATGATGATATACAAACATATATCTTTAGAAAGCATTTATCTGAAGAT 879
Oy 286 caattaataagaggtgtaagcttccgagaagaataatgaatcaactagatgatacaaatat 345
Db 880 AAGAAAAAATTAAGACATTTATGCGAAGATGCTTTATTCATGACATGATATGTATAT 939
Oy 346 tcaattgcagagcacacttcaatlgacgctccgaactatacaattacaacttcaaga 405
Db 940 GCAAAAGAAAGGATATGACACCCCTACTGTAATCCATCTTCGGAAGATTATAGAAAAT 999
Oy 406 gccaaaaaagtagtagtggtaacttaaaagtgtgaatgaacacgtaagataaaatg 465
Db 1000 ACTGAAAGGCGACTGACGTTTATGAAATAGATATATTCAGAGGATATTTTA 1059
Oy 466 acaagataagagatgtaagcctacagatgtaagagttctagat 510
Db 1060 AGTAAATTAATCAACCATATCAGAAATTTTATGATTAAT 1104

```

RESULT 15

US-08-082-849B-1
Sequence 1, Application US/08082849B
Patent No. 5677274

GENERAL INFORMATION:
APPLICANT: Leppla, Stephen H.
APPLICANT: Klimpel, Kurt R.
APPLICANT: Atora, Naveen
APPLICANT: Nichols, Peter J.
TITLE OF INVENTION: Anthrax Toxin Fusion Proteins and
NUMBER OF SEQUENCES: 35
CORRESPONDENCE ADDRESS:
ADDRESSEE: Townsend and Townsend and Crew LLP
STREET: Two Embarcadero Center, Eighth Floor
CITY: San Francisco
STATE: California
COUNTRY: USA

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/082.849B
FILING DATE: 25-JUN-1993
CLASSIFICATION: 514
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/021,601
FILING DATE: 12-FEB-1993
ATTORNEY/AGENT INFORMATION:
NAME: Weber, Kenneth A.

```

1  REGISTRATION NUMBER: 31,677
2  REFERENCE/DOCKET NUMBER: 15080-161-1
3  TELECOMMUNICATION INFORMATION:
4  TELEPHONE: (415) 576-0200
5  TELEFAX: (415) 576-0300
6  INFORMATION FOR SEQ ID NO: 1:
7  SEQUENCE CHARACTERISTICS:
8  LENGTH: 3291 base pairs
9  TYPE: nucleic acid
10 STRANDEDNESS: single
11 TOPOLOGY: linear
12 MOLECULE TYPE: DNA (genomic)
13 HYPOTHETICAL: NO
14 ANTI-SENSE: NO
15 ORIGINAL SOURCE:
16 ORGANISM: Bacillus anthracis
17 FEATURE:
18 NAME/KEY: CDS
19 LOCATION: 580..2907
20 OTHER INFORMATION: /product= "Lethal Factor"
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
841
842
843
844
845
846
847
848
849
850
851
852
853
854
855
856
857
858
859
860
861
862
863
864
865
866
867
868
869
870
871
872
873
874
875
876
877
878
879
880
881
882
883
884
885
886
887
888
889
890
891
892
893
894
895
896
897
898
899
900
901
902
903
904
905
906
907
908
909
910
911
912
913
914
915
916
917
918
919
920
921
922
923
924
925
926
927
928
929
930
931
932
933
934
935
936
937
938
939
940
941
942
943
944
945
946
947
948
949
950
951
952
953
954
955
956
957
958
959
960
961
962
963
964
965
966
967
968
969
970
971
972
973
974
975
976
977
978
979
980
981
982
983
984
985
986
987
988
989
990
991
992
993
994
995
996
997
998
999
1000
1001
1002
1003
1
```

Query Match	7.8%	Score 48.2;	DB 1;	Length 3291;
Best Local Similarity	44.9%	Pred. No. 0.004;		
Matches 182; Conservative	0;	Mismatches 223;	Indels 0;	Gaps 0;

Oy	106	aaagataaagatltgaaagatttgacaacatattgataaataatgataaagaacacgagtaa	165
Db	700	AAACNCATTTGTAAAAATATGAAAGTAAAAAGGGAGAAAGCTTTAAAAAAGAGCCACACAAA	759
Oy	166	gataaaagcagcaagcctaaccctcaattccgnaaagataaaatcgaaagtgcagctat	225
Db	760	AAGCTACTTGGAAGATGACCATCTGCATGTTTATAGATGATATAAAGCAATTTGAGCAAG	819
Oy	226	atlgaaattccagatgtcgtatattataaagaaccgatataccagacagcaacacctgaa	285
Db	820	ATATATTTGTGATGTGGTGAATTTAACAAAATATTTCTTTGAAAGCATTTATCTAAGAT	879
Oy	286	caatlaatagaagttgtaagctttgcagaaanaaigtacatcagatgataaataatt	345
Db	880	AAGAAAAAATAAAGACATTTTATGGGAAAGATGCTTTATATACATGAACATTATGTATAT	939
Oy	346	tcaattgcagagacacatttcatttgacggtccgaaactatacattcaaatctlaagca	405
Db	940	GCAAAAGAGAGATATGAAACCCGTACTTTGATATCCATCTTTGGAGAGATTATGACAAAT	999
Oy	406	gccaaaaagtagtatgtgtatcctttaaagtgtglaatgaataaacgtaaglaataaatg	465
Db	1000	ACTGAAAAAGCAGCTGAACGTTTATTTATGAAATATGATGACATTTTATCAAGGGATATTTTAA	1059
Oy	466	acaagtataagagttgtaagccttaagatgagtagttcttgat	510
Db	1060	AGTAAATTTATACACCATATTCAGAAATTTTATGATGTATTTAAAT	1104

Search completed: March 14, 2001, 14:28:04
Job time: 27600 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 14, 2001, 06:46:34 ; Search time 1314.13 Seconds
(without alignments)
2418.424 Million cell updates/sec

Title: US-09-292-437-2

Perfect score: 621
Sequence: 1 atgaaaaatggaacaatcg.....tagctacaagaagtaataaa 621

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1118133 segs, 2558875100 residues

Total number of hits satisfying chosen parameters: 2236266

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

GenEmbl:*

1: gb_da1:*
2: gb_da2:*
3: gb_om:*
4: gb_ov:*
5: gb_ph:*
6: gb_pi1:*
7: gb_pi2:*
8: gb_pi3:*
9: gb_pi4:*
10: gb_pi5:*
11: gb_pi6:*
12: gb_pi7:*
13: gb_pi8:*
14: gb_pi9:*
15: gb_pi10:*
16: gb_pi11:*
17: gb_pi12:*
18: gb_pi13:*
19: gb_pi14:*
20: gb_pi15:*
21: gb_pi16:*
22: gb_pi17:*
23: gb_pi18:*
24: gb_pi19:*
25: gb_pi20:*
26: gb_pi21:*
27: gb_pi22:*
28: gb_pi23:*
29: gb_pi24:*
30: gb_pi25:*
31: gb_pi26:*
32: gb_pi27:*
33: gb_pi28:*
34: gb_pi29:*
35: gb_pi30:*
36: gb_pi31:*
37: gb_pi32:*
38: gb_pi33:*
39: gb_pi34:*
40: gb_pi35:*
41: gb_pi36:*
42: gb_pi37:*
43: gb_pi38:*

44: em_htg2:*
45: em_htg3:*
46: em_hum5:*
47: gb_pi3:*
48: gb_pi5:*
49: gb_pi8:*
50: gb_pi9:*
51: gb_pi10:*
52: gb_pi11:*
53: gb_pi12:*
54: gb_pi13:*
55: gb_pi14:*
56: gb_pi15:*
57: gb_pi16:*
58: gb_pi17:*
59: gb_pi18:*
60: em_hum4:*
61: em_hum5:*
62: em_hum6:*
63: em_hum7:*
64: em_hum8:*
65: gb_pi18:*
66: gb_pi19:*
67: gb_pi20:*
68: gb_pi21:*
69: gb_pi22:*
70: gb_pi23:*
71: gb_pi24:*
72: gb_pi25:*
73: gb_pi26:*
74: em_hum8:*
75: em_hum9:*
76: em_hum10:*
77: gb_pi6:*
78: gb_pi7:*
79: gb_pi8:*
80: gb_pi9:*
81: gb_pi10:*
82: gb_pi11:*
83: em_hum9:*
84: gb_pi24:*
85: gb_pi28:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	621	100.0	1256	1	AF162687
2	340.2	54.8	2356	1	AF269739
3	315.2	50.8	4017	1	AF269386
4	72.4	11.7	152409	84	PFMALP1
5	68	11.0	196490	29	AC005507
6	67.2	10.8	2069	21	E10125
7	67.2	10.8	3399	21	E10126
8	66.8	10.8	1686	21	E08995
9	66.6	10.7	131271	40	AC015927
10	66.4	10.7	1324	56	PERESAR1
11	66.4	10.7	4591	81	A00661
12	65.4	10.5	7218	31	I66494
13	64.4	10.4	12029	31	AE001373
14	64.2	10.3	201289	49	AC021008
15	64	10.3	879	79	CNS01JRG
16	64	10.3	51014	39	AC011856
17	63	10.1	143585	39	AC013349
18	62.8	10.1	5420	32	AF056936
19	62.4	10.0	202748	84	CNS05TER
20	62	10.0	83440	51	AC024285
21	62	10.0	153477	29	AC006278

C	22	62	10.0	166050	51	AC024352	AC024352 Homo sapi
C	23	61.2	9.9	7852	31	AB018798	AB018798 Plasmodiu
C	24	60.6	9.8	11829	31	AE001376	AE001376 Plasmodiu
C	25	60.6	9.8	192581	84	PFMAL13P1	AE0019180 Plasmodiu
C	26	60.4	9.7	14760	31	AE001410	AE001410 Plasmodiu
C	27	60.2	9.7	156550	40	AC015830	AC015830 Homo sapi
C	28	60	9.7	825	32	AF206632	AF206632 Plasmodiu
	29	59.8	9.6	2349	81	A00764	A00764 Synthetic P
	30	59.8	9.6	4739	56	PFACIUPRA	M59706 P.falcipar
	31	59.8	9.6	192929	29	AC005505	AC005505 Plasmodiu
	32	59.4	9.6	153936	85	AC079863	AC079863 Homo sapi
C	33	59.2	9.5	170132	64	CNS05TDS	AL357093 Homo sap
C	34	59	9.5	840	7	CNS0180K	AL110675 Botrytis
C	35	59	9.5	74119	54	AC036177	AC036177 Homo sapi
C	36	59	9.5	113880	56	PFMAL3P4	AL008970 Plasmodiu
C	37	58.8	9.5	175957	40	AC015961	AC015961 Homo sapi
C	38	58.6	9.4	86932	50	AC022680	AC022680 Homo sapi
C	39	58.4	9.4	110000	84	PFMAL13P2_1	AC022680 Homo sapi
C	40	58.2	9.4	95764	40	AC015732	AC015732 Homo sap
C	41	58.2	9.4	200125	54	AC034105	AC034105 Homo sapi
C	42	57.8	9.3	81120	50	AC022851	AC022851 Homo sapi
C	43	57.8	9.3	164399	56	PFMAL3P6	298551 Plasmodiu
C	44	57.8	9.3	245802	29	AC006279	AC006279 Plasmodiu
C	45	57.4	9.2	3085	56	DD060169	U60169 Dictyosteli

ALIGNMENTS

Query	Match	Similarity	100.0%	Score 621	DB 1	Length 1256	
Best	Local	Similarity	100.0%	Pred. NO. 6.7e-89			
Matches	621	Conservative	0	Mismatches	0	Indels	0
						Gaps	0
QY	1	atgaaaaaatgacaaacgcgtatgaagaacgcgtggtgtaacttaacctagtgcga	60				
Db	483	ATGAAAAATGACAAATCGATTATGACAAATCGCTGGTGGTACTTATCTCAATGGCA	542				
QY	61	gcataattggttgcgtaaaccacatcgtaaatctctcacgataaagaataagtga	120				
Db	543	GCATTTTGTGTCGTAACACCATCGATTATTTATCTTCACGATAAAGATTAAGTCGA	602				
QY	121	aagcgttgaaacatatgataaanaattaaagaacggcgagtaagaataaanaagcagca	180				
Db	603	AAGATTGAACATATGATTAATAATATTAAGAAGACAGGCCAGTAAGATTAAGAAGCGCA	662				
QY	181	gctaaacctcaaatccgaaagaataaaccgaagtgcgcgtctatattgaaattccagat	240				
Db	663	GCTAAACCTCAAAATTCGGAACATTAATCGAAGTGGCAGCCTATATTGAAATTCCAGAT	722				
QY	241	gctgatattaaagaaccgatatatccaggaccagacaacctgaaacattaaatagagt	300				

RESULT	1
AF162687	
LOCUS	AF162687 1256 bp DNA BCT 11-AUG-1999
DEFINITION	Staphylococcus aureus sortase (srta) gene, complete cds.
ACCESSION	AF162687
VERSION	AF162687.1 GI:5726435
KEYWORDS	.
SOURCE	Staphylococcus aureus.
ORGANISM	Staphylococcus aureus Bacteria; Firmicutes; Bacillus/Clostridium group; Bacillus/Staphylococcus group; Staphylococcus. 1 (bases 1 to 1256) Mazmanian,S.K., Liu,G., Ton-That,H. and Schneewind,O. Staphylococcus aureus sortase, an enzyme that anchors surface proteins to the cell wall Proteins to the cell wall 285 (3428), 760-763 (1999)
JOURNAL	MEDLINE 9935/874
PUBMED	10427003
REFERENCE	2 (bases 1 to 1256)
AUTHORS	Mazmanian,S.K., Liu,G., Ton-That,H. and Schneewind,O.
TITLE	Direct Submission
JOURNAL	Submitted (24-JUN-1999) Microbiology and Immunology, UCLA, 10833 Le Conte Avenue, Los Angeles, CA 90095, USA
FEATURES	Location/Qualifiers

```
source
1. .1256
  /organism="Staphylococcus aureus"
  /strain="8325-4"
  /db_xref="taxon:1280"
483. .1103
  /gene="srta"
  /gene="srta"
483. .1103
CDS
```

BASE COUNT	470 a	185 c	234 g	367 t
ORIGIN				

RESULT	2
AF269739/c	
LOCUS	AF269739 2356 bp DNA BCT 01-AUG-2000
DEFINITION	Staphylococcus epidermidis strain SRI clone step.1020h12 genomic
ACCESSION	Sequence.
VERSION	AF269739
KEYWORDS	AF269739.1 GI:9623635
SOURCE	Staphylococcus epidermidis.

REFERENCE AUTHORS	TITLE	JOURNAL REFERENCE
1 (bases 1 to 2356) Kimmelman, W. J., Taylor, J. David, Nelsen, A. J., Godlewski, M. M., Rubino, M. A., Nelson, F. J., Rivers, P. R., Tortorella-Miller, I., Listebsee, S., Ashanti, C., Alshuler, G., Mam, L., Shepherd, N. S., Fuchs, R., Fleming, T., Guan, X., Du, L., Cain, D. H., Miller, G. S. and Furdon, P. J.	Transposon-mediated sequencing of the <i>Staphylococcus epidermidis</i> genome	2 (bases 1 to 2356)

AUTHORS Taylor, J. David, Kimmerly, W. J., Nelsen, A. J., Godlewski, M. M., Rubino, M. A., Nelson, F. J., Rivers, P. R., Tortorella-Miller, I., Listenebee, S., Ashanti, C., Altschuller, G., Mamo, L., Shepherd, N. S., Fuchs, R., Fleming, T., Guan, X., Du, L., Cain, D. H., Miller, G. S. and Furdon, P. J.

TITLE Direct Submission

JOURNAL Submitted (22-MAY-2000) Departments of Genomic Sciences and Bioinformatics, Genetics Directorate, Glaxo Wellcome, Inc., 5 Moore Drive, Research Triangle Park, North Carolina 27709-3398, USA

FEATURES Location/Qualifiers

source 1. 2356 /organism="Staphylococcus epidermidis" /strain="SRI" /db_xref="taxon:1282" /clone="step.1020h12"

BASE COUNT 777 a 371 c 332 g 876 t

ORIGIN

Query Match 54.8%; Score 340.2; DB 1; Length 2356;
Best Local Similarity 73.1%; Pred. No. 7.9e-45;
Matches 454; Conservative 0; Mismatches 158; Indels 9; Gaps 1;

1 atgaaataatgacaaatcgatgaatcgctgtgtgtctctatcctagtgcga 60
1222 ATGAAGCAGGTGATGAATAGATTAATCAGCTTAATAGCGGTATTGTTAATCAATTTAGCT 1163

61 gcaattgttgcctcaaacacatcgataatcttcacagataaagaataagatga 120
1162 ATTATTTATTTCTCAAGCCATATTCCTATTTATCTACATGAAGAAAGATACATATCAT 1103

121 aagattgacaatataatgaataatgtaaaagacagcgagtaagaatgaagaacagcaa 180
1102 AAAATTGAAATTTATGATTAATAAAGAAAAAGAACAGAACATCATCTA-----AA 1052

181 gctaactcaaatccgaaagataatcgaaagtggcagctatataatgcaatccagat 240
1051 TCGAGCGCAAAAGATACCTTCCTCAATTAATCGCTGTTATATAGCAACTTCCAGAT 992

241 gctgatatgaagaccagatatccagagcagacacacacccgaaatgaatagagat 300
991 GCACAAATTAAGAACACAGTATACCTGCTCCAGACACCCAGAACACATCATAGAGAT 932

301 gtaagccttcgaaagaataatgaatcactagatgaatcaaatatcttcagagacac 360
931 GTTAACTTTTGCGAAGGTGACGATCTCTTAATCAACAGATATTTCAATTCCTGCTCAT 872

361 acttcattgaccgctccgaactataatctcaaatcttaaaagcccaaaaagtgatg 420
871 ACGTTTACAGATCGTTCCGACTATCAATTTACAAATTTAAATTCAGCCAAATTCGCTACT 812

421 atggttaacttaagttgtaatgaacacgtaagtaaaatgacaagatagaagat 480
811 AAAGGTATTTTAAAGCTGGAATACAACTACAAAGTATTAATAATTAATTAATTAATTAAT 752

481 gtaagcctacagatgtagagttctagatgaacaaaagaagtaagaataaacaatga 540
751 GTTAAAGCCCTACAGAGGTATTAAGTATTAAGCAACATCTCTTAATTAAGAAAAATCAATTAACA 692

541 ttaattacttgatgattacaatgaaaaagcagcgcttgggaaaaaagtaaatcttt 600
691 TTTAATTTACTTCGATGACTATTAACAGAAACGGGTTTGGGAAACAAAGAAATATTTC 632

601 gtagctacagaagtaacaataa 621
631 ATAGCTACCAATTAATTAACATA 611

RESULT 3
AF269386 4017 bp DNA BCT 03-AUG-2000
LOCUS AF269386 Staphylococcus epidermidis strain SRI clone step.1003a02 genomic
DEFINITION sequence.

ACCESSION AF269386

VERSION AF269386.1 GI:9664707

KEYWORDS Staphylococcus epidermidis.

SOURCE Staphylococcus epidermidis

ORGANISM Bacteria; Firmicutes; Bacillus/Clostridium group; Staphylococcus.

REFERENCE 1 (bases 1 to 4017)

AUTHORS Kimmerly, W. J., Taylor, J. David, Nelsen, A. J., Godlewski, M. M., Rubino, M. A., Nelson, F. J., Rivers, P. R., Tortorella-Miller, I., Listenebee, S., Ashanti, C., Altschuller, G., Mamo, L., Shepherd, N. S., Fuchs, R., Fleming, T., Guan, X., Du, L., Cain, D. H., Miller, G. S. and Furdon, P. J.

TITLE Transposon-mediated sequencing of the Staphylococcus epidermidis genome

JOURNAL Unpublished

REFERENCE 2 (bases 1 to 4017)

AUTHORS Taylor, J. David, Kimmerly, W. J., Nelsen, A. J., Godlewski, M. M., Rubino, M. A., Nelson, F. J., Rivers, P. R., Tortorella-Miller, I., Listenebee, S., Ashanti, C., Altschuller, G., Mamo, L., Shepherd, N. S., Fuchs, R., Fleming, T., Guan, X., Du, L., Cain, D. H., Miller, G. S. and Furdon, P. J.

TITLE Direct Submission

JOURNAL Submitted (19-MAY-2000) Departments of Genomic Sciences and Bioinformatics, Genetics Directorate, Glaxo Wellcome, Inc., 5 Moore Drive, Research Triangle Park, North Carolina 27709-3398, USA

FEATURES Location/Qualifiers

source 1. 4017 /organism="Staphylococcus epidermidis" /strain="SRI" /db_xref="taxon:1282" /clone="step.1003a02"

BASE COUNT 1270 a 647 c 616 g 1484 t

ORIGIN

Query Match 50.8%; Score 315.2; DB 1; Length 4017;
Best Local Similarity 75.0%; Pred. No. 6.1e-41;
Matches 411; Conservative 0; Mismatches 126; Indels 9; Gaps 1;

74 ctaaacacatatcgataatattcttcacgataaagaataagaatgaatgaacat 133
4016 CAAGCCATATATCGATTAATTTATCATGAAAAAGATACATCAATTAATTTGAATTT 3957

134 atgataaaatgtaaaagacagcgagtaagaataaagaagcagaagctaaacctcaaa 193
3956 ATGATTAATAAGAAAAAGAACAGACAAACATCTTAA-----TCGACGCCAAAGA 3906

194 ttccgaagaataaatcgaaagtggcagctatattgtaaatcttcagatgctgatatgaag 253
3905 TACCTTCCGATTAATCTTAATTAATGCGTGTATATAGACATTCAGATCAAAATTAAG 3846

254 aaccgatataatccagacagcaaaccccgaaatgaatgaagtgtagagcttgag 313
3845 AACCAATATACCTCGTCCAGCAACACCAACCAACTCAATAGAGGTGTTAGTTTGCAG 3786

314 aagaataatgaatcactagatgaatcaaatatttcaatcttcagagacacattcatgacc 373
3785 AAGGTGACCAATCTCTTAATCAACAGATATTTCAATTCCTGCTCATTAACGATTC 3726

374 gtcgaactatacaatttacaatcttaaaagccagcaaaaagtgatgtagtacttta 433
3725 GTTCGCACTATCAATTTTACAAATTTAAATTCAGCCAAATTCGATTAAGGTATTTTA 3666

434 aagcttgatgaataacagtaagtaaaaatgacaagatgaagagatgtagagctagcag 493
3665 AAACCTGGAATCAAACTAATAAGTATTAATACTTAATTAACGTGATTTAAGCCCTACAG 3606

494 atgtagaagttctagatgaacaaaagtaagaataaacaatgaatgaatgaatcttgg 553
3605 AGCTTAAGGTATTTACAGCAACATCTTAATTAAGAAAAATCAATTAATTAATTAATTAAT 3546

554 atgattacaatgaagaagcagcgcttgggaaaaacgtaaaatctttagtagcagaag 613

Db	3545	ATGACTATTACGAAGAAACGGGTGTTTGGGAACAAGAAATATTCACTAGCTACACANA	3486
OY	614	tcaataa 621 	
Db	3485	TTAACATAA 3478	
RESULT	4		
PFMALIP1			
LOCUS			
DEFINITION		Plasmodium falciparum chromosome 1 strain 3D7, *** SEQUENCING IN PROGRESS ***, in unordered pieces.	
ACCESSION		AL031744	
VERSION		AL031744.7 GI:5731887	
SOURCE		HTG; HTGS-PHASEI.	
ORGANISM		malaria parasite P. falciparum.	
REFERENCE		Plasmodium falciparum	
AUTHORS		Eukaryota: Alveolata: Apicomplexa: Haemosporida: Plasmodium.	
TITLE		1 (bases 1 to 152409)	
JOURNAL		Bowman,S., Churcher,C., Harris,B., Harris,D., Lawson,D., Quail,M. and Barrell,B. Direct Submission Submitted (24-SEP-1998) P.falciparum Genome Sequencing Consortium, The Sanger Centre, Wellcome Trust Genome Campus, Hinxton, Cambridge CB10 1SA, UK On Aug 12, 1999 this sequence version replaced gi:5706481. For more information about this sequence or the Malaria Project, see http://www.sanger.ac.uk/Projects/P_falciparum. IMPORTANT: This sequence is unfinished and does not necessarily represent the correct sequence. Work on the sequence is in progress and the release of this data is based on the understanding that the sequence may change as work continues. The sequence may be contaminated with foreign sequence from E.coli,Yeast, vector, phage etc. Order of segments is not known: 800 n's separate segments. * NOTE: This is a 'working draft' sequence. * This record will be updated with the finished sequence * as soon as it is available and the accession number will * be preserved.	
FEATURES			
source		Location/Qualifiers 1..152409 /organism="Plasmodium falciparum" /strain="3D7" /db_xref="taxon:5833" /chromosome="1"	
BASE COUNT		52517 a 14209 c 13674 g 51206 t 20803 others	
ORIGIN			
Query Match		11.7%; Score 72.4; DB 84; Length 152409;	
Best Local Similarity		47.0%; Pred.No.0.0043;	
Matches	223;	Conservative 0; Mismatches 251; Indels 0; Gaps 0;	
OY	72	tgcttaaccacatcgtaatattccttcacgataagaataagatgaagaattgaaca 131 	
Db	151312	TGTAGAACAAAATGTAGAACAAAATGTAGAACAAAATGTAGAACAAAATGTGAGAAGAAA 151371	
OY	132	atafgataaaatgtaaagaacggcgagtgaagataaaaagccgcagcctaaccctca 191 	
Db	151372	TGTAGAACAAAATGTGTAAGAAAATGTAGAACAAAATGTAGAACAAAATGTGGAAGAAATA 151431	
OY	192	aattccgaagaataatcgaagaatgycagcgcatactgaattccagatgcgatatta 251 	
Db	151432	TGATCGAACAAAATGTGTGAAGAAGTAGAACAAAATGTGTGAAGAATGTGATGAGAAAAATGT 151491	
OY	252	agaaccgatlatatccaggaccagcaaaccttgaacaattaaatagaagtgttagacttgc 311 	
Db	151492	TGACAGCAAGTAGAACAAAATGTGTAGAACAAAATGTACAGCAAAATGTGTGAGAAAGAAATGTTTGA 151551	
OY	312	agaagaanaatgatcatcagatgatccaataattccaattgcyagcacacctttcaatga 371 	
Db	151552	AGAAATATGATGAGAAAATGTGTGAAGAAGTAGAACAAAATGTGTGAGAAAGAAATGTGAGAGA 151611	

QY	372	ccgtccgaccatccatcattaccacatccttaaaagcagcccaaaaggtgattggttact	431
Db	151612	AAATGTTGAAGAAAATGTGAGAAAGAAAATGTTCAAGAACTAGCAAGAAAATGTTGAGAGAAA	151671
QY	432	taaaagttgttaatgtgaacacgtlaaglatanaatgacaaaglataaagatgtttaagccttac	491
Db	151672	TGTAGTAAGAAAATGTGAGAAAGAAATGTTGAGAGAAATGTTGAGAGAAATGTTGAGAGAAATA	151731
QY	492	agatgttagagattcctagatgtaacaaaagtgtaagatatacaatacaattat	545
Db	151732	TGATGAGAAAATATGTTGAGAAACACAAATGAGAAATATGATGATATATATATATATATATAT	151785
RESULT	5		
LOCUS	AC005507/c		
DEFINITION	Plasmodium falciparum chromosome 12 clone 3D7, *** SEQUENCING IN PROGRESS ***. 2 unordered pieces.		
ACCESSION	AC005507		
VERSION	AC005507		
KEYWORDS	AC005507.8 GI:9797737		
SOURCE	HTG; HTGS_PHASE1		
ORGANISM	malaria parasite P. falciparum.		
TITLE	Plasmodium falciparum		
REFERENCE	Eukaryota: Alveolata: Apicomplexa: Haemosporida: Plasmodium.		
AUTHORS	1 (bases 1 to 196490) Hyman, R.W., Fung, E.L., Qin, F., Rowley, D., Mao, J., Tanaki, T., Kundi, O.B., Conway, A.B. and Davis, R.W.		
JOURNAL	Plasmodium falciparum 3D7 chromosome 12		
AUTHORS	Unpublished		
JOURNAL	2 (bases 1 to 196490) Hyman, R.W., Qin, F., Fung, E.L., Conway, A.B. and Davis, R.W.		
TITLE	Submitted (21-AUG-1998) Stanford DNA Sequencing and Technology Center, Stanford University, 855 California Avenue, Palo Alto, CA 94304, USA		
COMMENT	On Aug 12, 2000 this sequence version replaced gi:8810456. * NOTE: This is a 'working draft' sequence. It currently * consists of 2 contigs. The true order of the pieces * is not known and their order in this sequence record is * arbitrary. Gaps between the contigs are represented as * runs of N, but the exact sizes of the gaps are unknown. * This record will be updated with the finished sequence * as soon as it is available and the accession number will * be preserved. * 1 181303: contig of 181303 bp in length * 181304 181503: gap of unknown length * 181504 196490: contig of 14987 bp in length. Location/Qualifiers 1. 196490 /organism="Plasmodium falciparum" /db_xref="taxon:5833" /chromosome="12" /clone="PFYACB8-628" /clone="3D7"		
BASE COUNT	77103 a 20186 c 18680 g 80315 t 206 others		
ORIGIN			
Query Match	11.0%; Score 68; DB 29; Length 196490;		
Best Local Similarity	47.4%; Pred. No. 0.02;		
Matches 240; Conservative	0; Mismatches 260; Indels 6; Gaps 1;		
QY	70	tttgctaaacacatattcgataatattatcttcaacgtlaaagaataaagatgaaagttaa	129
Db	44795	TATGTTACAAATTCACAAAGATATTTTATGATCAAGCAATAGAAAATATACATATATCA	44736
QY	130	caatcatgataaanaatgttaagaacagcgcgatlaagaataaanaagcagcaagctaacct	189
Db	44735	AACGAGGAAACACACTTAATATATAATATCATATCAAAAGGTAAAAAATAATAGCAAAATATAT	44676
QY	190	caaatccgaaagataa-----atcgaaagtgcgagcgtatattgaaatccagatgct	243

D	b6	44675	AATAAAGTGATGATACCAAAATATTAAATATCCGAAGATTTATTATTAATAATGAAAAGTATT	44616
Oy	244	gatactaaagaacccagatcatalccaggaccagcaacctgaacaacltaaatagagtgla	303	
D	b6	44615	AATAATATGATTAATTAATTAATTTACCAGAACAAATTAAGAATATATCATCAGAAATGTAAA	44556
Oy	304	agccttcgcagaagaataatgaaactcactatgatcatcaaaatlctcaattcgaggaacaact	363	
D	b6	44555	AGTGAGGCTATTAAAAAGAAAAGCTGGAAATATGATATATTCATTAATTTATTAATTAAT	44496
Oy	364	tccaatggaccgylccgacctatccattttaaacaacttaagccagaccaaaaaggtagtatg	423	
D	b6	44495	TTCATTCGTAAAAAGATATATGTACATGTGTTATTTGTGCGTAGAAGAGAAAAGACTGAAAACG	44436
Oy	424	glvtactttaaagtgtgtaalgtgaacacgylaaglaaaaaatgacaaagtataaacatgaagatglt	483	
D	b6	44435	ATGAAGAGGAATGATTAATAATATGTGAAGATGCATAAAAGAAAACATGAAAGAGATTAAT	44376
Oy	484	aagcctacagatgtgtgaggtlcttgatgaacaaagaagtlaaagaataaacatatcaatta	543	
D	b6	44375	AAGAAAGAAATTTATTAATTAGCAAACAAGAGAGAGGTGATACATCAAAAAGAAAACAAA	44316
Oy	544	attacttgttagtatgatacaatgaana 569		
D	b6	44315	AATATGTCGGATTAATAAATAATTAATAA 44290	
R	E	RESULT	6	
I	D	E10125	standard; RNA; UNC; 2069 BP.	
A	C	E10125:		
S	V	E10125.1		
D	F	08-OCT-1997 (Rel. 52, Created)		
D	F	02-SEP-2000 (Rel. 65, Last updated, Version 2)		
D	E	DNA encoding an immunogenicity protein.		
K	M	JP 1995284392-A/1.		
O	S	unidentified		
X	X	unclassified.		
R	N	[1]		
R	P	1-2069		
R	A	Doi H., Nagakuchi Y., Tanaka Y., Fujisaki Y.;		
R	T	"GENE CLONE MANIFESTING CHICKEN LEUCOCYTOZOOM IMMUNOGENIC PROTEIN AND GENE		
R	L	RECOMBINANT VACCINE AGAINST CHICKEN LEUCOCYTOZOOM";		
R	L	Patent number JP1995284392-A/1, 31-OCT-1995.		
R	L	DOUBUTSUYOU SEIBUTSUGAKUTEKI SEIZAI KYOKAI, KITASATO INST:THE.		
C	C	OS Leucocytozon caulleryi		
C	C	PN JP 1995284392-A/1		
C	C	PD 31-OCT-1995		
C	C	PI 19-APR-1994 JP 1994080643		
C	C	PI DOI HIROHITO, NAGAKUCHI YOSHIO, TANAKA YOSHIO, FUJISAKI YUJIRO		
C	C	PC C12N15/09,A6IK39/015,C1P21/02;		
C	C	strandness: Double;		
C	C	topology: Linear;		
C	C	Key Location/Qualifiers		
C	C	EH source	1..2069	/organism="Leucocytozon caulleryi"
C	C	FT		/strain="shizuoka"
C	C	FT		/note="EcORI adaptor"
C	C	FT		/note="EcORI adaptor"
C	C	FT		/note="EcORI adaptor"
C	C	FT		/product="Immunogenicity protein"
C	C	XX		

[illegible]


```

* 2766 27865: gap of 100 bp
* 2786 28807: contig of 942 bp in length
* 28808 28907: gap of 100 bp
* 28908 29847: contig of 940 bp in length
* 29848 29947: gap of 100 bp
* 29948 30897: contig of 950 bp in length
* 30898 30997: gap of 100 bp
* 30998 31912: contig of 915 bp in length
* 31913 32012: gap of 100 bp
* 32013 32917: contig of 905 bp in length
* 32918 33017: gap of 100 bp
* 33018 33967: contig of 950 bp in length
* 33968 34067: gap of 100 bp
* 34068 35010: contig of 943 bp in length
* 35011 35110: gap of 100 bp
* 35111 36047: contig of 937 bp in length
* 36048 36147: gap of 100 bp
* 36148 37083: contig of 936 bp in length
* 37084 37183: gap of 100 bp
* 37184 38131: contig of 948 bp in length
* 38132 38231: gap of 100 bp
* 38232 39180: contig of 949 bp in length
* 39181 39280: gap of 100 bp
* 39281 40231: contig of 951 bp in length
* 40232 40331: gap of 100 bp
* 40332 41226: contig of 895 bp in length
* 41227 41326: gap of 100 bp
* 41327 42275: contig of 949 bp in length
* 42276 42375: gap of 100 bp
* 42376 43301: contig of 926 bp in length
* 43302 43401: gap of 100 bp
* 43402 44336: contig of 935 bp in length
* 44337 44436: gap of 100 bp
* 44437 45355: contig of 919 bp in length
* 45356 45455: gap of 100 bp
* 45456 46372: contig of 917 bp in length
* 46373 46472: gap of 100 bp
* 46473 47424: contig of 952 bp in length
* 47425 47524: gap of 100 bp
* 47525 48460: contig of 936 bp in length
* 48461 48560: gap of 100 bp
* 48561 49507: contig of 947 bp in length
* 49508 49607: gap of 100 bp
* 49608 50571: contig of 964 bp in length
* 50572 50671: gap of 100 bp
* 50672 51589: contig of 918 bp in length
* 51590 51689: gap of 100 bp
* 51690 52631: contig of 942 bp in length
* 52632 52731: gap of 100 bp
* 52732 53692: contig of 961 bp in length
* 53693 53792: gap of 100 bp
* 53793 54720: contig of 928 bp in length
* 54721 54820: gap of 100 bp
* 54821 55741: contig of 921 bp in length
* 55742 55841: gap of 100 bp
* 55842 56781: contig of 940 bp in length
* 56782 56881: gap of 100 bp
* 56882 57794: contig of 913 bp in length
* 57795 57894: gap of 100 bp
* 57895 58846: contig of 952 bp in length
* 58847 58946: gap of 100 bp
* 58947 59869: contig of 923 bp in length
* 59870 59969: gap of 100 bp
* 59970 60880: contig of 911 bp in length
* 60881 60980: gap of 100 bp
* 60981 61900: contig of 920 bp in length
* 61901 62000: gap of 100 bp
* 62001 62918: contig of 918 bp in length
* 62919 63018: gap of 100 bp
* 63019 63968: contig of 950 bp in length
* 63969 64068: gap of 100 bp
* 64069 64986: contig of 918 bp in length
* 64987 65086: gap of 100 bp

```

```

* 65087 66019: contig of 933 bp in length
* 66020 66119: gap of 100 bp
* 66120 67033: contig of 914 bp in length
* 67034 67133: gap of 100 bp
* 67134 68116: contig of 983 bp in length
* 68117 68216: gap of 100 bp
* 68217 69171: contig of 955 bp in length
* 69172 69271: gap of 100 bp
* 69272 70184: contig of 913 bp in length
* 70185 70284: gap of 100 bp
* 70285 71229: contig of 945 bp in length
* 71230 71329: gap of 100 bp
* 71330 72293: contig of 964 bp in length
* 72294 72393: gap of 100 bp
* 72394 73305: contig of 912 bp in length
* 73306 73405: gap of 100 bp
* 73406 74335: contig of 930 bp in length
* 74336 74435: gap of 100 bp

```

Query Match 10.7%; Score 66.6; DB 40; Length 131271;
 Best Local Similarity 36.4%; Pred. No. 0.036;
 Matches 198; Conservative 0; Mismatches 346; Indels 0; Gaps 0;

```

Qy 77 aaccacatcgtacataatcattcaccgtaaagataaagatgaagaattgaacatg 136
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 55664 AANNATATATAGAGATGATATTAATAAAAAAAAAAAAAAAAAAAAAA 55605

Qy 137 ataaagttaagaagacgcgcgtaaagataaagacgcgcgtacacacacac 196
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 55664 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 55545

Qy 197 cgaagataaatacgaagtcgcgcgtataltgaattccagatgcgtataaagac 256
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 55544 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 55485

Qy 257 cagatataccgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 316
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 55484 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 55425

Qy 317 aaatgaatcactagatcatcaaatattcattgcgcgcacacattgcgcgc 376
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 55424 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 55305

Qy 377 cgaactatcaatttacaatttaagcgcacaaagtagatgltgacttaag 436
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 55364 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 55305

Qy 437 ttgtaatgaacacgcgtgaagataaataatgcacgtataagatgtaagcctaag 496
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 55304 AANNNAANNNANNNNAANNNNAANNNNAANNNNAANNNNAANNNNAANNA 55245

Qy 497 taggagttcgtatgaacaaagtaagataaagataaacaattacattactgtgtg 556
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 55244 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 55185

Qy 557 attcaatgaagaacgcgcgttcgtgaaacgcgttaaatcttltgactcagaagca 616
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 55184 ANANNAANNAANNAANNNANNNANNNANNNANNNANNNANNNANNNANNA 55125

Qy 617 ata 620
   |||
Db 55124 CAAA 55121

```

RESULT 10
 LOCUS PFRESARI 1324 bp mRNA INV 16-FEB-1998
 DEFINITION P.falciparum FC27 Ag46 RESA mRNA for ring-infected erythrocyte
 surface antigen.
 ACCESSION X05182
 VERSION X05182.1 GI:9962
 KEYWORDS antigen; ring-infected erythrocyte surface antigen; surface
 antigen; tandem repeat.

[illegible]

Db	919	AGAAATGTTGGAACACACATGAAGAAATGATGATAAAAAAAT	966
Oy	498	aggagttctagatgagacaaaagtgaaagatgaacaaatcaat	545
Db	919	AGAAATGTTGGAACACACATGAAGAAATGATGATAAAAAAAT	966


```

* 10585 11303: contig of 719 bp in length
* 11304 11403: gap of 100 bp
* 11404 12111: contig of 708 bp in length
* 12112 12211: gap of 100 bp
* 12212 12923: contig of 712 bp in length
* 12924 13023: gap of 100 bp
* 13024 13737: contig of 714 bp in length
* 13738 13837: gap of 100 bp
* 13838 14547: contig of 710 bp in length
* 14548 14647: gap of 100 bp
* 14648 15337: contig of 690 bp in length
* 15338 15437: gap of 100 bp
* 15438 16151: contig of 714 bp in length
* 16152 16251: gap of 100 bp
* 16252 16957: contig of 706 bp in length
* 16958 17057: gap of 100 bp
* 17058 17730: contig of 673 bp in length
* 17731 17830: gap of 100 bp
* 17831 18512: contig of 682 bp in length
* 18513 18612: gap of 100 bp
* 18613 19313: contig of 701 bp in length
* 19314 19413: gap of 100 bp
* 19414 20051: contig of 638 bp in length
* 20052 20151: gap of 100 bp
* 20152 20872: contig of 721 bp in length
* 20873 20972: gap of 100 bp
* 20973 21700: contig of 728 bp in length
* 21701 21800: gap of 100 bp
* 21801 22526: contig of 726 bp in length
* 22527 22626: gap of 100 bp
* 22627 23342: contig of 716 bp in length
* 23343 23442: gap of 100 bp
* 23443 24164: contig of 722 bp in length
* 24165 24264: gap of 100 bp
* 24265 24983: contig of 719 bp in length
* 24984 25083: gap of 100 bp
* 25084 25793: contig of 710 bp in length
* 25794 25893: gap of 100 bp
* 25894 26593: contig of 700 bp in length
* 26594 26693: gap of 100 bp
* 26694 27384: contig of 691 bp in length
* 27385 27484: gap of 100 bp
* 27485 28207: contig of 723 bp in length
* 28208 28307: gap of 100 bp
* 28308 29013: contig of 706 bp in length
* 29014 29113: gap of 100 bp
* 29114 29830: contig of 717 bp in length
* 29831 29930: gap of 100 bp
* 29931 30642: contig of 712 bp in length
* 30643 30742: gap of 100 bp
* 30743 31450: contig of 708 bp in length
* 31451 31550: gap of 100 bp
* 31551 32269: contig of 719 bp in length
* 32270 32369: gap of 100 bp
* 32370 33071: contig of 702 bp in length
* 33072 33171: gap of 100 bp
* 33172 33875: contig of 704 bp in length
* 33876 33975: gap of 100 bp
* 33976 34649: contig of 674 bp in length
* 34650 34749: gap of 100 bp
* 34750 35466: contig of 717 bp in length
* 35467 35566: gap of 100 bp
* 35567 36288: contig of 722 bp in length
* 36289 37121: contig of 733 bp in length
* 37122 37221: gap of 100 bp
* 37222 37928: contig of 708 bp in length
* 37929 38029: gap of 100 bp
* 38030 38740: contig of 711 bp in length
* 38741 38840: gap of 100 bp
* 38841 39549: contig of 709 bp in length
* 39550 39649: gap of 100 bp
* 39650 40340: contig of 691 bp in length

```

```

* 40341 40440: gap of 100 bp
* 40441 41164: contig of 724 bp in length
* 41165 41264: gap of 100 bp
* 41265 41957: contig of 693 bp in length
* 41958 42057: gap of 100 bp
* 42058 42739: contig of 682 bp in length
* 42740 42839: gap of 100 bp
* 42840 43571: contig of 732 bp in length
* 43572 43671: gap of 100 bp
* 43672 44390: contig of 719 bp in length
* 44391 44490: gap of 100 bp
* 44491 45209: contig of 719 bp in length
* 45210 45309: gap of 100 bp
* 45310 46033: contig of 724 bp in length
* 46034 46133: gap of 100 bp
* 46134 46841: contig of 708 bp in length
* 46842 46941: gap of 100 bp
* 46942 47618: contig of 677 bp in length
* 47619 47718: gap of 100 bp
* 47719 48439: contig of 721 bp in length
* 48440 48539: gap of 100 bp
* 48540 49193: contig of 654 bp in length
* 49194 49293: gap of 100 bp
* 49294 49950: contig of 657 bp in length
* 49951 50050: gap of 100 bp
* 50051 50765: contig of 715 bp in length
* 50766 50865: gap of 100 bp
* 50866 51570: contig of 705 bp in length
* 51571 51670: gap of 100 bp
* 51671 52408: contig of 738 bp in length
* 52409 52508: gap of 100 bp
* 52509 53185: contig of 677 bp in length
* 53186 53285: gap of 100 bp
* 53286 54003: contig of 718 bp in length
* 54004 54103: gap of 100 bp
* 54104 54814: contig of 711 bp in length
* 54815 54914: gap of 100 bp
* 54915 55643: contig of 729 bp in length
* 55644 55743: gap of 100 bp
* 55744 56475: contig of 732 bp in length
* 56476 56575: gap of 100 bp
* 56576 57279: contig of 704 bp in length

Query Match      10.3%  Score 64.2; DB 49; Length 201289;
Best Local Similarity 39.0%; Pred.No. 0.08;
Matches 201; Conservative 0; Mismatches 315; Indels 0; Gaps 0;

Oy 106 aagaataagatgaagaatggaacaatataatgtaaaatgtaagaacagcgagtaaa 165
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 159388 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 159329

Oy 166 gataaaagcgcaagcctcaactccgaagaataatggaagtgcgagctat 225
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 159328 AAAAAAAAAAAGGANNNNNNAAAAAAAAAAAAAAAAAAAAAAAAAANNN 159269

Oy 226 attgaattccagatgctgatatataagaaccagtatataccgagcgcacactgaa 285
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 159268 AAAAAAAAAAANNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 159209

Oy 286 caatataatagagtgtaagcttgccagaagaataatgtaactgtagtcaaatctt 345
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 159208 NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 159149

Oy 346 tcaattgcagcagacacttcattgacgcctccgaactcaacttaacatcttaagca 405
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 159148 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 159089

Oy 406 gccaaaaaaggtagatgltgltacttaagttggtlaaagcaacgtaagtataaatg 465
      ||||| ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 159088 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 159029

Oy 466 acaagataagagatgttaagcctacagatgtaagagttcagatgtaacaaagtaaa 525
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

[illegible]

Query Match	10.3%	Score 64	DB 79	Length 879
Best Local Similarity	37.9%	Pred. NO. 0.24	Mismatches 280	Indels 0
Matches	202	Conservative 51	Mismatches 280	Indels 0
OY	89	ataattatcttcacgataaagataagaatgaaagatgaaacatatgataaagttaa	148	
Db	844	AAAAAAAAAAAAAAAAAAAAAAAAAGTACTAATATATTGTGMAATGWTAAATAAAAA	785	
OY	149	aagacacgcgagtaagataaaagcgacgactcaaacctcaagtctcgaaagataat	208	
Db	784	AAAAAAAAAAAAAAAAAGCTWAATTAATAAATAAATAATGTGAATAKMAAATAATGTAA	725	
OY	209	cgaagctgacgacttatctgaaattccagatgcatatataaagaacagatatccag	268	
Db	724	AAARATAGMGWADKKAKWMAAAGKTRRAKRCGMAAATAAATAAATAAAGKA	665	
OY	269	gacgcgacacccctgacaatataagatagatgtaagctctgcagaagaataatgatacc	328	
Db	664	ATTAATAAAAAAAAAAAAAAAAAAGGTTATTAKKKAKWAKMAAATAAATAKGAATAAT	605	
OY	329	tagatgatcaaaataattccaattgcagacacacttccaattgaccgtccgaactcaat	388	

[illegible]

Search completed: March 14, 2001, 14:30:10
Job time: 27816 sec

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 14, 2001, 02:37:09 ; Search time 975.78 Seconds
(without alignments)
4459.663 Million cell updates/sec

Title: US-09-292-437-2

Perfect score: 621

Sequence: 1 atgaaaaaatggacaatcg.....tagctacagaagtcacataa 621

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 7991742 segs, 3503743858 residues

Total number of hits satisfying chosen parameters: 15983484

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

EST:*

1: gb_est1:*
2: gb_est2:*
3: gb_est3:*
4: gb_est4:*
5: gb_est5:*
6: gb_est6:*
7: gb_est7:*
8: gb_est8:*
9: gb_est9:*
10: gb_est10:*
11: gb_est11:*
12: gb_est12:*
13: gb_est13:*
14: gb_est14:*
15: gb_est15:*
16: gb_est16:*
17: gb_est17:*
18: gb_est18:*
19: gb_est19:*
20: gb_est20:*
21: gb_est21:*
22: gb_est22:*
23: gb_est23:*
24: gb_est24:*
25: gb_est25:*
26: gb_est26:*
27: gb_est27:*
28: gb_est28:*
29: gb_est29:*
30: gb_est30:*
31: gb_est31:*
32: gb_est32:*
33: gb_est33:*
34: gb_est34:*
35: gb_est35:*
36: gb_est36:*
37: gb_est37:*
38: gb_est38:*
39: gb_est39:*
40: gb_est40:*
41: em_estba:*
42: em_estfun:*
43: em_esthum1:*

44: em_esthum2:*
45: em_esthum3:*
46: em_esthum4:*
47: em_esthum5:*
48: em_esthum6:*
49: em_esthum7:*
50: em_esthum8:*
51: em_esthum9:*
52: em_esthum10:*
53: em_esthum11:*
54: em_esthum12:*
55: em_esthum13:*
56: em_esthum14:*
57: em_esthum15:*
58: em_esthum16:*
59: em_esthum17:*
60: em_esthum18:*
61: em_esthum19:*
62: em_esthum20:*
63: em_estin1:*
64: em_estin2:*
65: em_estin3:*
66: em_estin4:*
67: em_estlov1:*
68: em_estlov2:*
69: em_estpl1:*
70: em_estpl2:*
71: em_estpl3:*
72: em_estpl4:*
73: em_estpl5:*
74: em_estro1:*
75: em_estro2:*
76: em_estro3:*
77: em_estro4:*
78: em_estro5:*
79: em_estro6:*
80: em_estro7:*
81: em_estro8:*
82: em_estro9:*
83: em_estro10:*
84: em_estro11:*
85: em_estro12:*
86: em_estro13:*
87: gb_esta1:*
88: gb_esta2:*
89: gb_esta3:*
90: gb_esta4:*
91: gb_esta5:*
92: gb_esta6:*
93: gb_esta7:*
94: gb_esta8:*
95: gb_esta9:*
96: gb_esta10:*
97: gb_esta11:*
98: gb_esta12:*
99: gb_esta13:*
100: gb_esta14:*
101: gb_esta15:*
102: gb_esta16:*
103: gb_esta17:*
104: gb_esta18:*
105: gb_esta19:*
106: gb_esta20:*
107: gb_esta21:*
108: gb_esta22:*
109: gb_esta23:*
110: gb_esta24:*
111: gb_esta25:*
112: em_esthum21:*
113: em_esthum22:*
114: em_esthum23:*
115: em_estom1:*
116: em_estom2:*

117: em_estp16:*
 118: em_estp17:*
 119: em_estp18:*
 120: em_estp14:*
 121: em_estp15:*
 122: em_estp16:*
 123: em_estp17:*
 124: em_estp18:*
 125: em_estp19:*
 126: gb_est58:*
 127: gb_est59:*
 128: gb_est60:*
 129: gb_est61:*
 130: gb_est62:*
 131: gb_est63:*
 132: gb_est64:*
 133: gb_est65:*
 134: gb_est66:*
 135: gb_est67:*
 136: gb_est68:*
 137: gb_est69:*
 138: gb_est70:*
 139: gb_est71:*
 140: gb_est72:*
 141: gb_est73:*
 142: gb_est74:*
 143: gb_est75:*
 144: gb_est76:*
 145: gb_est77:*
 146: gb_est78:*
 147: gb_est79:*
 148: gb_est80:*
 149: gb_est81:*
 150: gb_est82:*
 151: gb_est83:*
 152: gb_est84:*
 153: gb_est85:*
 154: gb_est86:*
 155: gb_est87:*
 156: gb_est88:*
 157: gb_est89:*
 158: gb_est90:*
 159: gb_est91:*
 160: gb_est92:*
 161: gb_est93:*
 162: gb_est94:*
 163: gb_est95:*
 164: gb_est96:*
 165: gb_est97:*
 166: gb_est98:*
 167: gb_est99:*
 168: gb_est100:*
 169: gb_est101:*
 170: gb_est102:*
 171: gb_est103:*
 172: gb_est104:*
 173: gb_est105:*
 174: gb_est106:*
 175: gb_est107:*
 176: gb_est108:*
 177: gb_est109:*
 178: gb_est110:*
 179: gb_est111:*
 180: gb_est112:*
 181: gb_est113:*
 182: gb_est114:*
 183: gb_est115:*
 184: gb_est116:*
 185: gb_est117:*
 186: gb_est118:*
 187: gb_est119:*
 188: gb_est120:*
 189: gb_est121:*

SUMMARIES

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

Result No.	Score	Query Match Length	DB ID	Description
1	74.2	11.9	942 190	CNS018GS
2	70.6	11.4	593 190	CNS000880
3	70.6	11.4	894 105	BE215316
4	70.6	11.4	997 190	CNS0134P
5	69.2	11.1	581 191	CNS010PZ
6	69.2	11.1	942 190	CNS00HP7
7	69	11.1	959 190	CNS00655
8	69	11.1	1223 183	B12981
9	68.8	11.1	639 190	CNS0170D
10	68.8	11.1	1101 190	CNS0153V
11	68.8	11.1	1135 191	CNS033G0
12	68.4	11.0	870 156	AQ330286
13	68.4	11.0	1101 190	CNS01807
14	67.4	10.9	952 172	AQ897460
15	67	10.8	1101 190	CNS00KHB
16	66.8	10.8	796 190	CNS0118D
17	66.6	10.7	1101 190	CNS000H9
18	66.2	10.7	850 190	CNS000E7
19	66	10.6	576 191	CNS035N7
20	66	10.6	963 190	CNS000A4L
21	65.6	10.6	776 190	CNS009BD
22	65.6	10.6	780 110	BE642964
23	65.4	10.5	834 183	B12387
24	65.4	10.5	1198 183	BE420745
25	65.4	10.5	1185 178	AZ196926
26	65.4	10.5	1185 178	AZ196926
27	65.2	10.5	770 176	AZ186328
28	65.2	10.5	850 176	AZ186328
29	65	10.5	890 148	AQ026918
30	65	10.5	1101 190	CNS012NV
31	64.8	10.4	1059 190	CNS0022B
32	64.4	10.4	614 190	CNS0152H
33	64.4	10.4	888 191	CNS027EX
34	64.4	10.4	1101 190	CNS001EB
35	64.4	10.4	1101 190	CNS0039R
36	64.4	10.4	1101 190	CNS017OV
37	64.2	10.3	730 171	AQ879158
38	64.2	10.3	741 190	CNS007YN
39	64.2	10.3	815 171	AQ853920
40	64.2	10.3	865 151	AQ324474
41	64.2	10.3	1101 190	CNS000JT
42	64	10.3	662 192	CNS0419K
43	64	10.3	693 169	AQ782441
44	64	10.3	1042 190	CNS0148K
45	63.8	10.3	736 190	CNS009DE

ALIGNMENTS

RESULT 1
 CNS018GS/c
 LOCUS
 DEFINITION
 CNS018GS 942 bp DNA
 Drosophila melanogaster genome survey sequence T7 end of BAC
 BACN13p09 of DrosBAC library from Drosophila melanogaster (fruit fly), genomic survey sequence.
 ACCESSION
 VERSION
 KEYWORDS
 AL109318 GI:5629622
 GSS
 26-JUL-1999

SOURCE	fruit fly.
ORGANISM	Plasmodium Drosophila melanogaster
REFERENCE	Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
AUTHORS	1 (bases 1 to 942)
TITLE	Genoscope.
JOURNAL	Direct Submission
COMMENT	Submitted (23-JUL-1999) Genoscope - Centre National de Sequencage ; BP 191 91006 EVRY cedex - FRANCE (E-mail : segreff@genoscope.cns.fr - Web : www.genoscope.cns.fr) Determination of this BAC-end sequence was carried out as part of a collaboration with the European Drosophila Genome Project (EDGP) - http://www.edgp.ebi.ac.uk/. This Drosophila melanogaster BAC library (Dros BAC) was made by Alain Billard at CEPH (Centre d'Etude du Polymorphisme Humain) with funding provided by a MRC project grant. The DNA was prepared from embryos by Alain Bucheton and Genevieve Payan. It has been constructed in the vector pBelosBAC11.
FEATURES	Location/Qualifiers
Source	1..942
	/organism="Drosophila melanogaster"
	/plasmid="pBelosBAC11"
	/db_xref="taxon:71227"
	/clone_lib="DrosBAC"
	/clone="BACN13P09"
	/note="end : 17"
BASE COUNT	155 a 77 c 70 g 494 t 146 others
ORIGIN	
Query Match	11.9% Score 74.2; DB 190; Length 942;
Best Local Similarity	38.0% Pred. No. 1.5e-06;
Matches 178; Conservative 52; Mismatches 238; Indels 0; Gaps 0;	
OY	107 aagataaagatgaagaagttgacacatatgtataaagaatgtaagaacagcgagtaag 166
DB	575 AACAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAACAAAAAATTGAGAACNAAAC 516
OY	167 ataaaagcagcaagcttaaacctcaatccgaaagataatcgaaagtgcagcgtata 226
DB	515 GAAMAAAAAAAAAAGCTTAAMMMWMAAACAGARARAAAAANNNNNANNNAAAA 456
OY	227 ttgaatcccgatgctcatatlaaagaacagatattccgagccagcaacctgaac 286
DB	455 NAAAAAAAAAAAAAAAACTRAAMRAAATATRAAAAAAAMAAACMRAARAAAAAAW 396
OY	287 aataatagaagtgtaagcttgcagaagaatagaatcactagatgatacaaatatt 346
DB	395 AACCGGTARAAWTTAAAMWTGTAACCRRAAAATTAATAAAAAAAAAAAWAMARTAA 336
OY	347 caattgcagagcacacttcaattgacgcgcgaactcaatcttaaatcctaagcag 406
DB	335 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAMAMAMWTMACCMMAAAAAAAAA 276
OY	407 ccaaaaagtgatgtagtcttaactttaaagttgtaatgaacacgtaagtataatga 466
DB	275 AAAAAAAAAAAAAAAAAAMAAACMAAAAMWMAAAWMAAAWMTTAAWTTAAAAAAM 216
OY	467 caagataagaagtgtaagccttaagaagtgtagaggtctgatagaacaaaaagtaag 526
DB	215 WTGAAMAMAMWTMGCTTAAMAAATCGTTAMCAAAAMAMAAWMAAAWMAAAATWMAAA 156
OY	527 ataaacataacattacttacttgatgtaatgaatgaacgaagaagcag 574
DB	155 AMGCAAAATTTGATTTWAAAAAMAMAMAMWMAAAWMAAAWMAAAAG 108
RESULT 2	
CNS00880	593 bp DNA GSS 03-JUN-1999
LOCUS	Drosophila melanogaster genome survey sequence TE13 end of BAC #
DEFINITION	BACR16J23 of RFL1-98 library from Drosophila melanogaster (fruit

Accession	Version	Source	Organism	Reference	Author	Title	Journal
AL051540	1	GI:4933381	fruit fly.	Drosophila melanogaster	Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephyridiidae; Drosophilidae; Drosophila.	1 (bases 1 to 593)	Genoscope.
Submitted (02-JUN-1999)	Genoscope - Centre National de Sequencage	Bp 191 91006 EVR cedex - FRANCE (E-mail : seqref@genoscope.cns.fr	- Web : www.genoscope.cns.fr)	Determination of this BAC-end sequence was carried out as part of a collaboration with the Berkeley Drosophila Genome Project (BDGP).	The BDGP is constructing a physical map of the Drosophila melanogaster genome using these BACs. For further information please see http://www.fruitfly.org	The BDGP Drosophila melanogaster BAC library was prepared by Kazuhiro Osoegawa and Aaron Mamoser in Pieter de Jong's laboratory in the Department of Cancer Genetics at the Roswell Park Cancer Institute in Buffalo, NY. The library is named RPCI-98 and was constructed by partial EcoRI digestion of Drosophila DNA provided by the BDGP from the isogenic strain Y2; cn bw sp, the same strain used for the BDGP's P1 and EST libraries. A more detailed description of the library and how to order individual BAC clones, the entire library, or filters for hybridization from the BACPAC Resource Center can be found at http://bacpac.med.buffalo.edu/drosophila_bac.htm .	Location/Qualifiers
1..593	/organism="Drosophila melanogaster"	/db_xref="taxon:7227"	/clone_lib="RPCI-98"	/clone="BACR16J23"	/note="end : TET3"		
BASE COUNT	448 a	25 c	18 g	10 t	92 others		
ORIGIN							
Query Match	11.4%	Score 70.6:	DB 190;	Length 593;			
Best Local Similarity	37.9%	Pred. No. 9e-06;					
Matches 209;	Conservative 56;	Mismatches 286;	Indels 0;	Gaps 0;			
QY	71	ttgtctaaaccacatcatcgcataatcttcacgcgtaaagataaagatgaagatggaac	130				
Db	6	TGCDCAAAATACAAAGGAAAAAAATAATCAACATTAATAATTAACACMGCAAAAAAACMAA	65				
QY	131	aataatgataaaaatgtaaaagaacagcgagtgtaaaagataaaagcagaactaacctc	190				
Db	66	AAAAAAMAAAAAMWMAAAAAAAMAAAAAAMAAATTAACMAAMMMCAAMMAAAAAA	125				
QY	191	aaatccgaagaatgaatcgaagaatgagcgagctatcgtgaatccagaatcgtgata	250				
Db	126	MAAGACGACAAAAMAAAMMAACAAAAAAMAAAAAAMAAAAAAMAAAAAAMAAAAA	185				
QY	251	aagaacagatataccagagccagcaacacactgaaacaaataaataagatgtaagcttg	310				
Db	186	AAAAAAMAAAAAMWMAAAAAAAMAAAAAAMAAATTAACMAAMMMCAAMMAAAAAA	245				
QY	311	cagaagaataatgaatcgaagaatgataatcgaataatcgaatcgcgcagacacttcattg	370				
Db	246	AGAACRMAAAAAAAMAAAAAAMAAAAAAMAAAAAAMAAAAAAMAAAAAAMAAAAA	305				
QY	371	acggtccgaactatcaattcaaatcttaaaagcagcacaagaagtgatgtgtact	430				
Db	306	AAMMAAAMAAAAAAMAAAAAAMAAAAAAMAAAAAAMAAAAAAMAAAAAAMAAAAA	365				
QY	431	ttaaatgtggttaaatgaaacacgtaagtataaatgacagatataagagatgtaagcta	490				
Db	366	ARAAAAAAMAAAAAAMAAAAAAMAAAAAAMAAAAAAMAAAAAAMAAAAAAMAAAAA	425				

[illegible]

Db	376	AAATTAATAAAAAAAAAAATAATAATAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA	435
Oy	284	aacaattaatagagggttaagcttgcagaagaatgaatcactgatatcaaaata	343
Db	436	AAAATTAAAAAATAACCAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAATA	495
Oy	344	tttcaattgcagggcacaccttcattgcgcgtccgaactacaaatltaacaatctaaag	403
Db	496	ATTAATAAAAAAAAAAAAAAAAAAATAATAATAATAATAATAAAAAAAAAAAAAA	555
Oy	404	cagccaataaaagtagtatggttgtaactttaagtgtgtaacgaaacgctaataaa	463
Db	556	AAATAATAATAATAATAA	615
Oy	464	tgacagatataagatgttgttaagcctaagaatgttagsggtcttgatgtaaaaaaagta	523
Db	616	AAAAAAAAAAAAATAATAATAAAAAAAAAAAAAAAAAATAAAAAAAAAAAAAAAAAA	675
Oy	524	aagataaacataatacataacttactgtgtgatatacatgaaaagcgagcgtttgg	583
Db	676	AATTAATAATAATAATAATAATAATAATAATAATAAAAAAAAAAAAAAAAAAAAA	735
Oy	584	aaaaacgttaaat 596	
Db	736	AAAAAAAAAAAAAT 748	
RESULT	4	CNS0134P/c	GSS
LOCUS		CNS0134P	997 bp DNA GSS 26-JUL-1999
DEFINITION		Drosophila melanogaster genome survey sequence Sfe end of BAC BACN09C07 of DrosBAC library from Drosophila melanogaster (fruit fly), genomic survey sequence.	
ACCESSION		AL102403	
VERSION		ALI02403.1	GI:5614014
KEYWORDS		GSS.	
SOURCE		fruit fly.	
ORGANISM		Plasmid Drosophila melanogaster Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscoroidea; Ephydroidea; Drosophilidae; Drosophila. 1 (bases 1 to 997)	
REFERENCE		Direct Submission Submitted (23-JUL-1999) Genoscope - Centre National de Sequencage : BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr Web : www.genoscope.cns.fr) Determination of this BAC-end sequence was carried out as part of a collaboration with the European Drosophila Genome Project (EDGP) - http://www.edgp.ebi.ac.uk/. This Drosophila melanogaster BAC library (Dros BAC) was made by Alain Billaud at CEPH (Centre d'Etude du Polymorphisme Humain) with funding provided by a MRC project grant. The DNA was prepared from embryos by Alain Bucheton and Genevieve Payan. It has been constructed in the vector pBelobAC11.	
COMMENT			
TITLE	JOURNAL		
AUTHORS			
FEATURES			
Source		location/Qualifiers	
		1..997	
		/organism="Drosophila melanogaster"	
		/plasmid="pBelobAC11"	
		/db_xref="taxon:7227"	
		/clone_lib="DrosBAC"	
		/clone="BACN09C07"	
		/note="end : SP6"	
BASE COUNT	162 a 115 c 79 g 517 t 124 others		
ORIGIN			
Query Match	11.4%; Score 70.6; DB 190; Length 997;		
Best Local Similarity	39.1%; Pred. No. 9.3e-06;		
Matches	185; Conservative 42; Mismatches 243; Indels 3; Gaps 1;		
Oy	76	aaaccacatcogatattatcttcagcataaagaatgaaaagattgaaacatat	135
	:	: :	:

please see <http://www.fruitfly.org> The BGP Drosophila melanogaster BAC library was prepared by Kazutoyo Osoegawa and Aaron Mammosier in Pleier de Jong's laboratory in the Department of Cancer Genetics at the Roswell Park Cancer Institute in Buffalo, NY. The library is named RPCI-98 and was constructed by partial EcoRI digestion of Drosophila DNA provided by the BGP from the isogenic strain y2; cn bw sp, the same strain used for the BGP's pl and EST libraries. A more detailed description of the library and how to order individual BAC clones, the entire library, or filters for hybridization from the BACPAC Resource Center can be found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.

FEATURES

source
1. 942
/organism="Drosophila melanogaster"
/db_xref="taxon:7227"
/clone_lib="RPCI-98"
/clone="BACR35109"
/note="end : 17"
BASE COUNT 459 a 62 c 76 g 144 t 201 others
ORIGIN

Query Match 11.1%; Score 69.2; DB 190; Length 942;
Best Local Similarity 35.7%; Pred. No. 1.9e-05;
Matches 168; Conservative 69; Mismatches 233; Indels 0; Gaps 0;

QY 77 aaccacatcgaataatcttcacgataaagaataagaatgaacaataatg 136
+ :
Db 342 AATCAAAAGAAAAAATGACAAAAAAGAAAAAAGAAAAAAGAAAAAAG 401
+ :
QY 137 ataaaaatgaaagaacgagagtaaaaaagaacgaacgaacgaacgaac 196
+ :
Db 402 AAAAAAAGAAAAAAGAAAAAAGAAAAAAGAAAAAAGAAAAAAGAAAA 461
+ :
QY 197 cgaagaataacgaagcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 256
+ :
Db 462 AATCAAAAGAAAAAATGACAAAAAAGAAAAAAGAAAAAAGAAAAAAG 521
+ :
QY 257 cagatataccagacgaacacgcgcgcgcgcgcgcgcgcgcgcgcgcgc 316
+ :
Db 522 SAACBSYTTABWRTTCTTAAMHMTASABKKAAYTAAWYAWGATHTWT 581
+ :
QY 317 aaaaatgaatcagatgataaataatcttcacatgacgacacacacacac 376
+ :
Db 582 ASAAATTTATATAMGCTAAAGCTGAAAGBGTATGSMTTAAATTAAMGMA 641
+ :
QY 377 cgaactacatcttaacatcttaacgaacgaacgaacgaacgaacgaac 436
+ :
Db 642 TAAMGCTAAGTCAKATATCCCAAAAAAASAAACAAKATTAATATGATG 701
+ :
QY 437 ttgtaataagaacgtaagataaataagaacgaacgaacgaacgaacga 496
+ :
Db 702 GGAATTTDTMTTAAATGATTAACMSACASATGTAATTAATAAADDMD 761
+ :
QY 497 taggaattcagatgaacaaaggttaagaataaacaataacataac 546
+ :
Db 762 GATKAGATAGAGTGAARAARAGRTWAVATWASAAATTTWGSAGAT 811
+ :
RESULT 7
CNS00655 959 bp DNA GSS 03-JUN-1999
LOCUS Drosophila melanogaster genome survey sequence TET3 end of BAC #
DEFINITION BACR13F18 of RPCI-98 library from Drosophila melanogaster (fruit
fly), genomic survey sequence.
ACCESSION AL062806 GI:4943996
VERSION
KEYWORDS
SOURCE fruit fly.
ORGANISM Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.

REFERENCE 1 (bases 1 to 959)
AUTHORS Genoscope.
TITLE Direct Submission
JOURNAL Submitted (02-JUN-1999) Genoscope - Centre National de Sequenage ;
BP 191 91006 Evry cedex - FRANCE (E-mail : seque@genoscope.cns.fr
- Web : www.genoscope.cns.fr)

COMMENT

Determination of this BAC-end sequence was carried out as part of a collaboration with the Berkeley Drosophila Genome Project (BDGP). The BGP is constructing a physical map of the Drosophila melanogaster genome using these BACs. For further information please see <http://www.fruitfly.org> The BGP Drosophila melanogaster BAC library was prepared by Kazutoyo Osoegawa and Aaron Mammosier in Pleier de Jong's laboratory in the Department of Cancer Genetics at the Roswell Park Cancer Institute in Buffalo, NY. The library is named RPCI-98 and was constructed by partial EcoRI digestion of Drosophila DNA provided by the BGP from the isogenic strain y2; cn bw sp, the same strain used for the BGP's pl and EST libraries. A more detailed description of the library and how to order individual BAC clones, the entire library, or filters for hybridization from the BACPAC Resource Center can be found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.

FEATURES

source
1. 959
/organism="Drosophila melanogaster"
/db_xref="taxon:7227"
/clone_lib="RPCI-98"
/clone="BACR13F18"
/note="end : TET3"
BASE COUNT 525 a 63 c 52 g 106 t 213 others
ORIGIN

Query Match 11.1%; Score 69; DB 190; Length 959;
Best Local Similarity 36.6%; Pred. No. 2.1e-05;
Matches 190; Conservative 89; Mismatches 225; Indels 5; Gaps 2;

QY 101 acgaataagaatgaagaatgaacaaatgaataaagaacgaacga 160
+ :
Db 426 AAAAAAAGAAAAAATTTTGGTGGTGTGAAATGAAAGAAAGAAAGAA 485
+ :
QY 161 gtaagaataaagaacgaacgaacgaacgaacgaacgaacgaacga 220
+ :
Db 486 AAAAAAAGAAAAAAGAAAAAAGAAAAAAGAAAAAAGAAAAAAGAAAA 545
+ :
QY 221 gctatgaatccgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 280
+ :
Db 546 AARGGGAARAAAAAATTAATAAAAAAATAAATAAATAAATAAATAA 604
+ :
QY 281 ctgaacaataatagagtgtaagcttcgagaagaataatgaatcagatga 340
+ :
Db 605 --FAAAGCAAAWMAKADAAAAAADAAGCAKAAATATATTAATAA 661
+ :
QY 341 atattcgaatgacgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 400
+ :
Db 662 WTTTATTATTTW -AKRAAAAAAATAATADWDMAAAWMAAAAAAADA 720
+ :
QY 401 aagcagcgaacgaacgaacgaacgaacgaacgaacgaacgaacgaacga 460
+ :
Db 721 AARAAAAAAGAAAAAAGAAAAAAGAAAAAAGAAAAAAGAAAAAAG 780
+ :
QY 461 aaatgaacgataagaatgaacgaacgaacgaacgaacgaacgaacga 520
+ :
Db 781 GDTTAAARAAKRRKCPAAAAAATTAARAAAAAADAADKAAAAAAGAA 840
+ :
QY 521 gtaagaataacatcttaacatcttaacatcttaacatcttaacatct 580
+ :
Db 841 RAAAAAAGAAAAAAGAAAAAAGAAAAAAGAAAAAAGAAAAAAGAA 900
+ :
QY 581 ggggaacgaacgaacgaacgaacgaacgaacgaacgaacgaacgaac 619
+ :
Db 901 GKRAAAAAAAGAAAAAAGAAAAAAGAAAAAAGAAAAAAGAAATTAAT 939
+ :
RESULT 7
CNS00655 959 bp DNA GSS 03-JUN-1999
LOCUS Drosophila melanogaster genome survey sequence TET3 end of BAC #
DEFINITION BACR13F18 of RPCI-98 library from Drosophila melanogaster (fruit
fly), genomic survey sequence.
ACCESSION AL062806 GI:4943996
VERSION
KEYWORDS
SOURCE fruit fly.
ORGANISM Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.

```

RESULT      8
Bi2981/c    1223 bp    DNA    GSS    14-MAY-1997
LOCUS       T24D11-sp6 Arabidopsis thaliana genomic clone T24D11, DNA
DEFINITION  sequence.
ACCESSION   B12981
VERSION     B12981.1 GI:2094103
KEYWORDS    GSS.
SOURCE      thale cress.
ORGANISM    Arabidopsis thaliana
             Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
             Magnoliophyta; eudicotyledons; core eudicot; Rosidae; eurosids II;
             Brassicales; Brassicaceae; Arabidopsids.
REFERENCE   1 (bases 1 to 1223)
AUTHORS     Feng,J., Dewar,K., Buehler,E., Kim,C., Li,Y., Shinn,P., Sun,H. and
             Ecker,J.
TITLE       BAC End Sequences at ATGC
JOURNAL     Unpublished (1997)
COMMENT     Contact: Ecker J.
             Arabidopsis Thaliana Genome Center
             University of Pennsylvania
             Dept. of Biology, University of Pennsylvania, Philadelphia, PA
             19104
             Tel: 215-898-9384
             Fax: 215-898-8780
             Email: jecker@genome.bio.upenn.edu
             Seq primer: SP6
             Class: BAC ends
             High quality sequence start: 199
             High quality sequence stop: 271.
FEATURES
  source     1..1223
             /organism="Arabidopsis thaliana"
             /strain="Columbia"
             /db_xref="taxon:3702"
             /clone="T24D11"
             /clone_11b="TAMU"
             /sex="hermaphrodite"
             /note="Vector: BelobAC11; Site_1: HindIII; Site_2: HindIII
             ; Produced by Rod Wing"
BASE COUNT   70 a    50 c    53 g    678 t    372 others
ORIGIN
Query Match      11.1%; Score 69; DB 183; Length 1223;
Best Local Similarity 36.9%; Pred. No. 2.1e-05;
Matches 201; Conservative 0; Mismatches 344; Indels 0; Gaps 0;

Qy 76 aaacacatcgtataatcttcaagataagataagatgaagaattgacaatat 135
    || || || || || || || || || || || || || || || || || ||
Db 1145 AANNNAANNAANNAANNAANNAANNAANNAANNAANNAANNAANNAAN 1086

Qy 136 gataaataatgtaagaacagcgagtaaagataaagaacagcctaactcaatt 195
    || || || || || || || || || || || || || || || || || ||
Db 1085 NAAANNAANNAANNAANNAANNAANNAANNAANNAANNAANNAANNA 1026

Qy 196 ccgaagaataaactcgaagctgcaagctatattgaattccagatgctgataaaga 255
    || || || || || || || || || || || || || || || || || ||
Db 1025 AANNAANNAANNAANNAANNAANNAANNAANNAANNAANNAANNAANNA 966

Qy 256 ccagtatatccgagcagcagcaacacgtgaacaattaaatagagtgtaagcttgcaga 315
    || || || || || || || || || || || || || || || || || ||
Db 965 AAAAAANNAANNAANNAANNAANNAANNAANNAANNAANNAANNAANNN 906

Qy 316 gaaatagaatcactagatgataaataattcgaattgcaggaacacttcatgtgcgt 375
    || || || || || || || || || || || || || || || || || ||
Db 905 AAAAAANNAANNAANNAANNAANNAANNAANNAANNAANNAANNAANNA 846

Qy 376 ccgaactatcaatttcaactcttaagcagccaaagaagtgatgtgtgtaacttaaa 435
    || || || || || || || || || || || || || || || || || ||
Db 845 AAAAAANNAANNAANNAANNAANNAANNAANNAANNAANNAANNAANNA 786

Qy 436 gtgtgtaatgaacacgtaagataaaatgacaagataagagatgtaagcctacagat 495

```

```

Db 785 AAAAAANNAANNAANNAANNAANNAANNAANNAANNAANNAANNAANNA 726
Qy 496 gtgaggtcttagatgacacaagaagtaagaataaacaattacttactgtgt 555
    || || || || || || || || || || || || || || || || || ||
Db 725 AAAAAANNAANNAANNAANNAANNAANNAANNAANNAANNAANNAANNA 666

Qy 556 gattacaatgaagaacagcgcttcttggaataaactcttgtagtcagagaatc 615
    || || || || || || || || || || || || || || || || || ||
Db 665 NAAAAANNAANNAANNAANNAANNAANNAANNAANNAANNAANNAANNN 606

Qy 616 aaata 620
    || ||
Db 605 NAANA 601

RESULT      9
CNS0170D    639 bp    DNA    GSS    26-JUL-1999
LOCUS       Drosophila melanogaster genome survey sequence SP6 end of BAC
DEFINITION  BAC37M13 of DrosBAC library from Drosophila melanogaster (fruit
             fly), genomic survey sequence.
ACCESSION   AL108367
VERSION     AL108367.1 GI:5628671
KEYWORDS    GSS.
SOURCE      fruit fly.
ORGANISM    Plasmid Drosophila melanogaster
             Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
             Pterygota; Neoptera; Endopterygota; Diptera; Brachyceta;
             Muscophora; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1 (bases 1 to 639)
AUTHORS     Genoscope.
TITLE       Direct Submission
JOURNAL     Submitted (23-JUL-1999) Genoscope - Centre National de Sequencage :
             BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
             - Web : www.genoscope.cns.fr)
             Determination of this BAC-end sequence was carried out as part of a
             collaboration with the European Drosophila Genome Project (EDGP) -
             http://www.edgp.ebi.ac.uk - This Drosophila melanogaster BAC
             library (Dros BAC) was made by Alain Billard at CERH (Centre
             d'Etude du Polymorphisme Humain) with funding provided by a MRC
             Project grant. The DNA was prepared from embryos by Alain Bucheton
             and Genevieve Payan. It has been constructed in the vector
             pBelobAC11.
FEATURES
  source     1..639
             /organism="Drosophila melanogaster"
             /plasmid="pBelobAC11"
             /db_xref="taxon:7227"
             /clone_11b="DrosBAC"
             /clone="BAC37M13"
             /note="end : SP6"
BASE COUNT   389 a    53 c    46 g    52 t    99 others
ORIGIN
Query Match      11.1%; Score 68.8; DB 190; Length 639;
Best Local Similarity 38.3%; Pred. No. 2.2e-05;
Matches 185; Conservative 51; Mismatches 247; Indels 0; Gaps 0;

Qy 91 aattatcttcagataaagaataagaatgaagaatgaacaatatgataaaatgtcaaa 150
    || || || || || || || || || || || || || || || || || ||
Db 140 AAAAAATTGAMARARACAAAMAAAYAAAHMRAMWMTWMAAAAAAGAAAAAATRAAAAAARAA 199

Qy 151 gaacagcgagtaagaataaagaacagcagcgaactaaactcgaagaatcgt 210
    || || || || || || || || || || || || || || || || || ||
Db 200 GTAGGAGGATTAATTAAMWTAAMRAAAMAAAMAAAMATATCAACWMAAAARFAATTA 259

Qy 211 aaagtgcgagcatttgaattccagatgctgatatgaagaacagatataccaaga 270
    || || || || || || || || || || || || || || || || || ||
Db 260 ARAAMWATGTGACTAAWMAAAWATWTAAMWRTTTTAAGAAAAAAMWGTGAWMAR 319

Qy 271 ccagcaacacctgaacaattaaatagagtgtaagcttgcagaagaataatgaactacta 330

```



```

/dbc_xref="taxon:99883"
/cdb="208P24"
/cclone.lib="c"
/note="Genoscope sequence ID : C0AC208BH12SP1-end
PUC-Or1"
BASE COUNT      863 a      65 c      43 g      124 t      40 others
ORIGIN

```

Query Match	11.1%	Score 68.8;	DB 191;	Length 1135;
Best Local Similarity	45.4%;	Pred No. 2.3e-05;		
Matches 235, Conservative	3;	Mismatches 280;	Indels 0;	Gaps 0

[illegible]

RESULT	12
LOCUS	A0330286/c
DEFINITION	A0330286 870 bp DNA GSS 08-JAN-1999
ACCESSION	nxbdb0066j18r CUGI Rice BAC Library Oryza sativa genomic clone
VERSION	nxbdb0066j18r, DNA sequence.
KEYWORDS	A0330286 A0330286.1 GI:4122136 GSS.
SOURCE	Oryza sativa.
ORGANISM	Oryza sativa Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Oryza. 1 (bases 1 to 870) Wing, R.A. and Dean, R.A. A BAC End Sequencing Framework to Sequence the Rice Genome Unpublished (1998) Contact: Wing RA Clemson University Genomics Institute Clemson University 100 Jordan Hall, Clemson, SC 29634, USA Tel: 864 656 7286 Fax: 864 656 4293 Email: rtwing@clemson.edu

```
Seq primer: GGAACACAGCTATGACACATC
Class: BAC ends
High quality sequence start: 13
High quality sequence stop: 104
location/Qualifiers
    source
        1..870
```

/lab host="E. coli DH10B"
 /note="Vector: pBlotBAC11; site_1: HindIII; site_2:
 HindIII. Rice is one of two most popular grains in the
 world. Half of the world population especially those
 inhabiting highly populated areas of the humid tropics
 and subtropics, rely on rice as their primary source of
 carbohydrate. Monocytyledonous rice is a diploid plant
 (2n=24) with a haploid genome equivalent of 431 Mbp
 (Armuganathan and Earle, 1991). The relatively small
 genome of rice, three times larger than that of
 Arabidopsis, makes it suitable for genomic studies. In
 order to facilitate positional cloning, physical mapping
 and genome sequencing of rice, we have constructed a BAC
 library from *Oryza sativa*, Nipponbare variety. The
 library contains 36,864 clones with an average insert size
 of 128.5 kb providing 10.9 haploid genome equivalents. The
 deep coverage allows the isolation a particular sequence
 with a probability of 99.9 %. Two high density filters,
 each containing 18,432 clones (doubly spotted), represent
 the whole library for colony screening."

Query Match	11.0%;	Score 68.4;	DB 156;	Length 870;
Best Local Similarly	42.3%;	Pred. No. 2.8e-05;		
Matches 219; Conservative	0;	Mismatches 299;	Indels 0;	Gaps 0;

[illegible]

```

QY 584 aaaaacgtaaatcttctagctcagaagtcacaataa 621
    ||| ||| | | | | | | | | | | | | | | | |
Db 283 NAAAAANNNAAAAANAAAAANAAAAANAAAAANAAAA 246

RESULT 13
LOCUS   CNS01807/c 1101 bp DNA GSS 26-JUL-1999
DEFINITION Drosophila melanogaster genome survey sequence sp6 end of BAC
            BACN37F07 of DrosBAC library from Drosophila melanogaster (fruit
            fly), genomic survey sequence.
ACCESSION AL108721
VERSION   AL108721.1 GI:5629025
KEYWORDS  GSS.
SOURCE    fruit fly.
ORGANISM  Plasmid Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
            Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
            Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE 1 (bases 1 to 1101)
AUTHORS   Genoscope.
TITLE     Direct Submission
JOURNAL   Submitted (23-JUL-1999) Genoscope - Centre National de Sequencage :
            BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
            - Web : www.genoscope.cns.fr)
            Determination of this BAC-end sequence was carried out as part of a
            collaboration with the European Drosophila Genome Project (EDGP) -
            http://www.edgp.ebi.ac.uk/. This Drosophila melanogaster BAC
            library (Dros BAC) was made by Alain Billaud at CEPH (Centre
            d'Etude du Polymorphisme Humain) with funding provided by a MRC
            project grant. The DNA was prepared from embryos by Alain Bucheton
            and Genevieve Payan. It has been constructed in the vector
            pBelobAC11.

FEATURES
    source             1..1101
                        /organism="Drosophila melanogaster"
                        /plasmid="pBelobAC11"
                        /db_xref="taxon:7227"
                        /clone_lib="DrosBAC"
                        /clone="BACN37F07"
                        /note="end : SP6"
BASE COUNT  179 a 70 c 85 g 448 t 319 others
ORIGIN
Query Match 11.0%: Score 68.4; DB 190; Length 1101;
Best Local Similarity 40.3%: Pred. No. 2.8e-05;
Matches 198: Conservative 25; Mismatches 268; Indels 0; Gaps 0;

QY 77 aaccacatcgaataatctatctcagataaagataaagatgaagaatg 136
    || | | | | | | | | | | | | | | | | | | | |
Db 580 AAAAAAAMWAAAAATATMAAAATMAAAAMWAAAAATMAAAAMWAAAAATMAA 521

QY 137 ataataatgtaagaagcagcgtaagaataaagacagcaagctaaacctcaatc 196
    ||| ||| | | | | | | | | | | | | | | | |
Db 520 AAAAAACAAAAAATMAAAATMAAAAMWAAAAAATMAAAATMAAAAMWAAAAA 461

QY 197 cgaagataaactcgaagtcgagctatctgaattccagatcgatataaagac 256
    ||| ||| | | | | | | | | | | | | | | | |
Db 460 AAAAAAAMWAAAAAATMAAAATMAAAAMWAAAAAATMAAAAMWAAAAA 401

QY 257 cagatatccagagcagcaacacctgaacaataatagagggtgaagcttgcagaag 316
    || | | | | | | | | | | | | | | | | | | | |
Db 400 AAAAAAAMWAAAAAATMAAAATMAAAAMWAAAAAATMAAAAMWAAAAA 341

QY 317 aaatgaatccctgagatgaataatctcaattgcagcgacacattcattgcgcgc 376
    ||| ||| | | | | | | | | | | | | | | | |
Db 340 AAAAAATATATAAAAAATATATMAWAAAAAATTAATATCAANNCAANTCTTCNCNA 281

QY 377 cgaactcaatttacaactctaaagcagcaaaaggtagctatgtgtaactttaaag 436
    || | | | | | | | | | | | | | | | | | | | |
Db 280 NCTGCCAATCATCAAAATTTATTTAAATTAATTAATTAATTAATTAATTAATTAAT 221

```

```

QY 437 ttgtaatgaaacacgctgaatataaalgacaagataaagatgttaagccctacagatg 496
    || | | | | | | | | | | | | | | | | | | | |
Db 220 AATATMAWAAAAAAMWAAAAATMAAAATNNANNNTAATTAATCTTAANNANNAATATN 161

QY 497 taggagctctgagatgaacaaaggtaagaataaacataactaattctctgtgctg 556
    || | | | | | | | | | | | | | | | | | | | |
Db 160 NNNAAAAAAMWAAAAAATMAAAATMAAAAMWAAAAAATMAAAATTAATTAATTTTNT 101

QY 557 attacaatgaa 567
    || | | | | | | | | | | | | | | | | | | | |
Db 100 ATTNNAAAAA 90

RESULT 14
LOCUS   AO897460 952 bp DNA GSS 10-NOV-1999
DEFINITION HS_3134_A2_G06_T7C CIT Approved Human Genomic Sperm Library D Homo
            sapiens genomic clone Plate=3134 Col=12 Row=M, DNA sequence.
ACCESSION AO897460
VERSION   AO897460.1 GI:6353650
KEYWORDS  GSS.
SOURCE    human.
ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 952)
AUTHORS   Mahaitas G.G., Wallace J.C., Smith R., Swartzell S., Holzman T.,
            Keller A., Shaker R., Furlong J., Young J., Zhao S., Adams M.D. and
            Hood L.
            Sequence-tagged connectors: A sequence approach to mapping and
            scanning the human genome
            Proc. Natl. Acad. Sci. U. S. A. 96 (17), 9739-9744 (1999)
JOURNAL   99380589
MEDLINE   Contact: Mahaitas GG, Wallace JC, Hood L
            High Throughput Sequencing Center
            University of Washington
            401 Queen Anne Avenue North, Seattle, WA 98109, USA
            Tel: (206) 616-3618
            Fax: (206) 616-3887
            Email: jwallace@u.washington.edu
            Clones may be purchased from Research Genetics (info@resgen.com).
            BAC end Web Server: http://www.hnsc.washington.edu
            Plate: 3134 row: M column: 12
            Seq primer: T7
            Class: BAC ends
            High quality sequence stop: 952.

FEATURES
    source             1..952
                        /organism="Homo sapiens"
                        /db_xref="taxon:9606"
                        /clone="Plate=3134 Col=12 Row=M"
                        /clone_lib="CIT Approved Human Genomic Sperm Library D"
                        /sex="male"
                        /note="Organ: sperm; Vector: pBelobAC11; BAC Clones in
                        E-Coli DH10B"
BASE COUNT  720 a 20 c 110 g 32 t 70 others
ORIGIN
Query Match 10.9%: Score 67.4; DB 172; Length 952;
Best Local Similarity 47.2%: Pred. No. 4.6e-05;
Matches 233: Conservative 0; Mismatches 259; Indels 2; Gaps 1;

QY 104 ataagataaagatgaagaatgtgaacatatgtataaagataaagagcgagta 163
    ||| ||| | | | | | | | | | | | | | | | |
Db 349 AGAANNAAAAAAMWAAAAACAAAAAAGAAAAAAGAAAAAAGAAAAAAGAA 408

QY 164 aagataaagcagcagctaaacctcaattccgaaagataaactcgaagctgagcagct 223
    || | | | | | | | | | | | | | | | | | | | |
Db 409 AAAAAAAMWAAAAAAMWAAAAAAMWAAAAAAMWAAAAAAMWAAAAAAMWAAAAA 468

QY 224 atattgaatccagatgctgatatlaaagaccagatattccaggaccagacacctg 283
    || | | | | | | | | | | | | | | | | | | | |

```


[illegible]

RESULT	15
CNS00KHB/c	
LOCUS	CNS00KHB 1101 bp DNA
DEFINITION	Drosophila melanogaster genome survey sequence T7 end of BAC: BACR17J10 of RPc1-98 library from Drosophila melanogaster (fruit fly), genomic survey sequence.

ACCESSION	AL077776	GI:457152
VERSION	AL077776.1	
KEYWORDS	GSF.	
SOURCE	fruit fly,	
ORGANISM	<i>Drosophila melanogaster</i>	
	Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta	
	Pterygota; Neoptera; Endopterygota; Diptera; Brachyera;	
	Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.	
REFERENCE	1 (bases 1 to 1101)	

REFERENCE
AUTHORS
TITLE
JOURNAL

COMMENT

Submitted (02-JUN-1999) Genoscope - Centre National de Séquençage
BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
Web : www.genoscope.cns.fr)
determination of this BAC-end sequence was carried out as part of a
collaboration with the Berkeley Drosophila Genome Project (BDGP).
The BDGP is constructing a physical map of the Drosophila
melanogaster genome using these BACs. For further information
please see <http://www.fruitfly.org> The BDGP Drosophila
melanogaster BAC library was prepared by Kazuhiro Oseegawa and
Aaron Mammos in Pieret de Jong's laboratory in the Department of
Cancer Genetics at the Roswell Park Cancer Institute in Buffalo,
NY. The library is named RPCT-98 and was constructed by partial
EcoRI digestion of Drosophila DNA provided by the BDGP from the
isogenic strain y2; cn bw sp. The same strain used for the BDGP's
P1 and EST libraries. A more detailed description of the library
and how to order individual BAC clones, the entire library, or
filters for hybridization from the BACpac Resource center can be
found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.

FEATURES
source

1. .1101

```
/organism="Drosophila melanogaster"
```

```
/db_xref=taxon:7227
```

```
/clone_id= "KPC175"
/clone= "BACR17J10"
```

```
/note="end : T7"
```

BASE COUNT	170 a	142 c	122 g	590 t	77 others
ORIGIN					

Query Match	10.8%;	Score 67;	DB 190;	Length 1101;
Best Local Similarity	42.4%;	Pred. No. 5.7e-05;		
Matches 211;	Conservative 23;	Mismatches 264;	Indels 0;	Gaps 0;

QY	76	aaaccaatattcgataattatcttccgaataagaataagatgaaagattgaaacaatat	133
Db	501	AA	44
QY	136	gataaataatgtaaaagacagcgtaagaataaaacagcagctaaacctcaatc	195
Db	441	AAATGAAACCAAMAA	38
QY	196	ccgaaagaataatcgaaatgctgcagctatacttgaaattccagatgcgatataaaga	255
Db	381	MAAAAAAAAAACMMAACAAACCAAAAAAAAAAAAAAAAAAAAAAAAAAATTAATTAATAA	32
QY	256	ccagatattccagagccagcaacacctgaaacaattaaataagatgtaagctgtgcaga	311
Db	321	AAGAAAAAAAAAAAAAAAAAGAAACCMAAAAAAAAAAAAAAAAACMAACAAAMWRAAAAAAAAA	26
QY	316	gaataatgatctactgatatgataataatttcaatttgcagagacacacttcaatgacgt	37
Db	261	MAAAAAAAAAMAATTAATAATCAAAAAAAAAAACCAACMAATTAATAAAAAAAAAAAAAA	20
QY	376	ccgaaactcaattaccaaacttaaaagcagccaaanaagtagatagtgtacttaaa	43
Db	201	CAAMTAAAAAAAAAAAAAAAAACAAAGAAAGAAAGAGTTGTRATTAAGAAAAAAAAAA	14
QY	436	gttggtaatgaacacgttaagtataaaatgacaagtataagatglttaagctcaagat	49
Db	141	AAAAAAAAAAAAAAAAAAAAAAAAAMTAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA	82
QY	496	gttagagttctagatgaaacaaaaggtaaagataaacaattaaacttaactgtgat	55
Db	81	AAAAAAAAAAWGAIAAAAAAAAAAAATTAATAAAAAAAAAAATTAATAAACCAAA	22
QY	556	gattcaatgaaagatca 573	
Db	21	AAAAAAAAAAAAAAAAAAAA 4	

Search completed: March 14, 2001, 14:04:33
Job time: 41244 sec

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 2, 2001, 10:22:33 ; Search time 9.87 Seconds
(without alignments)
674.021 Million cell updates/sec

Title: US-09-292-437-3
Sequence: 1 MKKWTNRLMTAGVLLVLA.....YNEKTVWEKKRIFVATEVK 206

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 88757 seqs, 32294092 residues
Total number of hits satisfying chosen parameters: 88757

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SwissProt_39.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	97.5	9.1	678	1 GARP_PLAIF	P13816 plasmodium
2	95.5	8.9	644	1 NFM_RABIT	P54338 oryctolagus
3	94.5	8.8	993	1 YIS2_YEAST	P40562 saccharomyc
4	91.5	8.5	198	1 NFM_RAR	P12839 rattus norv
5	91	8.5	198	1 YHCS_BACSU	P54603 bacillus su
6	91	8.5	1290	1 XCPC_XENLA	P50532 xenopus lae
7	90	8.4	915	1 NFM_HUMAN	P07197 homo sapien
8	89.5	8.3	411	1 FKB3_YEAST	P38811 saccharomyc
9	89	8.3	359	1 PMSR_HELPY	P025011 hellicobacte
10	89	8.3	472	1 YWIE_CAEEL	P08553 caenorhabdi
11	88	8.2	448	1 NFM_MOUSE	P25240 herpes simp
12	87	8.1	2077	1 TEGU_LISV6U	P50189 drosophila
13	85	7.9	922	1 TOP1_DROME	P56187 drosophila
14	84.5	7.9	195	1 YACB_BRELA	P54397 drosophila
15	84.5	7.9	357	1 FKB4_DROME	P54397 drosophila
16	84	7.8	2077	1 TEGU_HSV6G	P30002 herpes simp
17	84	7.8	6359	1 BACC_BACLI	P06808 b bacitraci
18	83.5	7.8	590	1 VNC3_ADVG	P24030 alectian mi
19	83.5	7.8	680	1 NASP_RABIT	P27123 oryctolagus
20	83.5	7.8	944	1 NUF1_YEAST	P32380 saccharomyc
21	83	7.7	218	1 YTXE_BACME	P46627 bacillus me
22	83	7.7	1080	1 SET1_YEAST	P38827 saccharomyc
23	82.5	7.7	448	1 PRTE_BACNO	P19577 bacteroides
24	82.5	7.7	527	1 G19P_HUMAN	P14314 homo sapien
25	82.5	7.7	845	1 SCPL_MESAU	P06363 mesocricetu
26	82	7.6	240	1 YPBE_BACSU	P50731 bacillus su
27	82	7.6	439	1 NFM_HUMAN	P55081 homo sapien
28	82	7.6	1020	1 NFM_HUMAN	P12036 homo sapien
29	82	7.6	1897	1 RPTF_HUMAN	P10586 homo sapien
30	81.5	7.6	827	1 RED1_YEAST	P14591 saccharomyc
31	81.5	7.6	967	1 SYL_PYRHO	O58698 pyrococcus
32	81	7.6	1002	1 IF2P_YEAST	P39730 saccharomyc
33	81	7.5	533	1 G19P_BOVIN	Q28034 bos taurus

34	81	7.5	1790	1 USO1_YEAST	P25386 saccharomyc
35	80.5	7.5	2468	1 MAPB_HUMAN	P46821 homo sapien
36	80	7.4	257	1 YZG1_CAEEL	P55326 caenorhabdi
37	80	7.4	273	1 OSAB_BORBU	Q09088 borrelia bu
38	80	7.4	457	1 CGPB_FUSSO	Q00858 fusarium so
39	80	7.4	735	1 FCTL_ERWCH	Q47162 erwilia chr
40	80	7.4	968	1 CC28_SCHPO	Q10752 schizosacch
41	80	7.4	2805	1 MAPA_HUMAN	P78559 homo sapien
42	79.5	7.4	424	1 IHBA_CHICK	P27092 gallus gall
43	79.5	7.4	1332	1 SPT7_YEAST	P35177 saccharomyc
44	79.5	7.4	1522	1 DNA2_YEAST	P38859 saccharomyc
45	79.5	7.4	1587	1 TOP2_PENCH	Q9y898 penicillium

ALIGNMENTS

RESULT 1	GARP_PLAIF	STANDARD;	PRT;	678 AA.
ID	GARP_PLAIF			
AC	P13816;			
DT	01-JAN-1990 (Rel. 13, Created)			
DT	01-JAN-1990 (Rel. 13, Last sequence update)			
DT	15-JUL-1999 (Rel. 38, Last annotation update)			
DE	GLUTAMIC ACID-RICH PROTEIN PRECURSOR.			
GN	GARP.			
OS	Plasmodium falciparum (isolate FC27 / Papua New Guinea).			
OC	Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.			
RN	(1)			
RP	SEQUENCE FROM N.A.			
RX	MEDLINE-89040048; PubMed-2903445;			
RA	Trigilia T., Stahl H.-D., Crewther P.E., Silva A., Anders R.F.,			
RA	Kemp D.J.;			
RT	"Structure of a Plasmodium falciparum gene that encodes a glutamic			
RT	acid-rich protein (GARP).";			
RL	Mol. Biochem. Parasitol. 31:199-202(1988).			
CC	-----			
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration			
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -			
CC	the European Bioinformatics Institute. There are no restrictions on its			
CC	use by non-profit institutions as long as its content is in no way			
CC	modified and this statement is not removed. Usage by and for commercial			
CC	entities requires a license agreement (See http://www.isb-sib.ch/announce/			
CC	or send an email to license@lsb-sib.ch).			
CC	-----			
DR	EMBL: J03998; AAA29605.1; -			
DR	PIR: A54514; A54514.			
KW	Repeat; Malaria; Antigen; Signal.			
FT	SIGNAL 1 25			
FT	CHAIN 26 678			
FT	DOMAIN 120 164			
FT	DOMAIN 372 416			
FT	DOMAIN 417 441			
FT	DOMAIN 576 604			
FT	DOMAIN 605 653			
FT	DOMAIN 654 663			
FT	SEQUENCE 678 AA; 80551 MW; 2A8F85606496EA9E CRC64;			
QY	Query Match	9.1%;	Score 97.5;	DB 1; Length 678;
QY	Best Local Similarity	24.1%;	Pred. No. 2.6;	
QY	Matches 42; Conservative	32;	Mismatches 73; Indels 27; Gaps 8;	
QY	128 HKRDKDEKIKYQKKNKEQSKKKQKQKQIKRDKSKVAGYIFIPADIKPEYFEPAT 93			
QY	128 HKRDKDEKIKYQKKNKEQSKKKQKQKQIKRDKSKVAGYIFIPADIKPEYFEPAT 93			
QY	94 PEOLNRCVSAEEN---ESLDONISITAGTFTIDRPYOPTNLKAAGKSGMYFKVQNET 150			
QY	181 PK---NATEIGENLDEEMYSFINNNAOGILLSP-YQYREBGGCGITISVH-ETNSNDT 235			
QY	151 KKKYKMSIRDKPPTD-----VGVLDQKQKDKROLTLITCDDYNEKTVWEKKR 198			

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: X96983; CA65703.1; -
 DR EMBL: 299108; CAB12748.1; -
 DR SUPPLIST: BG11597; Yhcs.
 KW Hypothetical protein: Transmembrane.
 FT TRANSMEM 7
 SQ SEQUENCE 198 AA: 21976 MW: 286C6F96CA96A43 CRC64;

Query Match 8.5%; Score 91; DB 1; Length 198;
 Best Local Similarity 23.4%; Pred. No. 1.8;
 Matches 51; Conservative 33; Mismatches 100; Indels 34; Gaps 10;

OY 1 MKKWTNRLMTIAGVLLIIVAAFLFAKPHIDNYLHDKDKDEIKDYDKNVK-----EQAS 54
 DB 1 MKKVIPLFIAGLVYAGYGFKLI-----DTWTKTEQTLKEAKLAKKPQAS 49
 OY 55 --KDKKQQA-KPQIPKSKVAGYIEIPDADIKEPVYGPATPEQNLNGVSFAENESL 110
 DB 50 GTRNSTDQAKNKASRPETGQASGILEIPKINAELEPVEG-TDADDELEKGVGHYKDSYYP 108
 OY 111 DQI-NISIAGHFTIDRPNQFTNLKAAKGSN--VYEKVGNEFRKYKMTSIDVPTDVG 167
 DB 109 DENGQIVLSGR-----DIVFRRTGELEKGDQLRLISGTEETTELVTIKYIVDKDDTSII 163
 OY 168 VIDEQKGDQQLTLITCDDYNEKTVGEKRIKIFVATEV 205
 DB 164 TLQHEK---EELILTYCPFS-YVGNAPKRYIIYGRV 197

RESULT 6

XPCPC_XENLA STANDARD; PRT; 1290 AA.

AC P50532;
 DT 01-OCT-1996 (Rel. 34, Created)
 DT 01-OCT-1996 (Rel. 34, Last sequence update)
 DT 15-JUL-1998 (Rel. 36, Last annotation update)
 DE CHROMOSOME ASSEMBLY PROTEIN XCAP-C.
 GN XCAP-C.
 OS Xenopus laevis (African clawed frog).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Amphibia; Batrachia; Anura; Mesobatrachia; Pipidoidea; Pipidae;
 OC Xenopodinae; Xenopus.
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=95042742; PubMed=7954811;
 RA Hirano T., Mitchison T.J.;
 RT "A heterodimeric coiled-coil protein required for mitotic chromosome
 condensation in vitro."
 RL Cell 79:449-458(1994).
 CC -1- FUNCTION: REQUIRED FOR BOTH ASSEMBLY AND STRUCTURAL MAINTENANCE OF
 MITOTIC CHROMOSOMES.
 CC -1- SUBUNIT: ASSOCIATES WITH XCAP-E PROBABLY AS HETERODIMER.
 CC -1- SUBCELLULAR LOCATION: NUCLEAR. DURING CHROMOSOME ASSEMBLY IN
 MITOTIC EXTRACTS, XCAP-C/E WAS RECRUITED TO THE CHROMATIN AND
 CC FORMED A DISCRETE INTERNAL STRUCTURE WITHIN ASSEMBLED CHROMOSOMES.
 CC -1- DOMAIN: CONSISTS OF TWO PUTATIVE CENTRAL COILED-COIL REGIONS
 CC FLANKED BY PUTATIVE GLOBULAR REGIONS AT THE N- AND C-TERMINUS.
 CC -1- SIMILARITY: BELONGS TO THE SMC FAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----

CC -----
 DR EMBL: U13673; AA64679.1; -
 KW Mitosis; ATP-binding; Coiled coil; Nuclear protein.
 FT NP BIND 107 114
 FT DOMAIN 107 114
 FT DOMAIN 264 594
 FT DOMAIN 764 1027
 FT DOMAIN 1094 1129
 FT DOMAIN 1263 1290
 FT DOMAIN 60 65
 FT DOMAIN 747 750
 FT DOMAIN 841 844
 FT DOMAIN 1196 1220
 FT ALA/ASP-RICH (DA-BOX).
 SQ SEQUENCE 1290 AA: 146988 MW: 2931249924FE90F6 CRC64;

Query Match 8.5%; Score 91; DB 1; Length 1290;
 Best Local Similarity 20.3%; Pred. No. 18;
 Matches 36; Conservative 43; Mismatches 60; Indels 38; Gaps 8;

OY 32 YLHD-----KQKD---EKIEQYKNNKEQA-----SKDKQQAQKQIPKDKSVAGYIE- 77
 DB 320 YIHDIQKRSRQEAQKEKIQEDFKDISKSNLTETMEKKNALKVDKQLKITFTIEE 379
 OY 78 -----IPDADIKEPVYGPATPEQNLNGVSFAENESLDD-QNISIAGHFTIDRPN 127
 DB 380 NNEKFTQLDQVDYDREKLKSKSVKRLQK--QLQDKREKVEDELKNVPANSQKIIAEET 437
 OY 128 YQFTLKAAGKGSWYFVNGNETRKYKMTSIRD-VKPTDVCYLDQKQKDKQQLTIT 183
 DB 438 NKKDLERQK-----EKEEKLNWDSLKTETGQLQEKKEKELMEIS 483

RESULT 7

NFM_HUMAN STANDARD; PRT; 915 AA.

AC P07197;
 DT 01-APR-1988 (Rel. 07, Created)
 DT 01-AUG-1988 (Rel. 08, Last sequence update)
 DT 01-OCT-2000 (Rel. 40, Last annotation update)
 DE NEUROFILAMENT TRIPLET M PROTEIN (160 KDA NEUROFILAMENT PROTEIN)
 DE (NEUROFILAMENT MEDIUM POLYPEPTIDE) (NF-M) (NEUROFILAMENT 3).
 GN NFM OR NFM OR NEF3.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=87275853; PubMed=3608989;
 RA Myers M.W., Lazzarini R.A., Lee V.M.-Y., Schlaepfer W.W., Nelson D.L.;
 RT "The human mid-size neurofilament subunit: a repeated protein
 RT sequence and the relationship of its gene to the intermediate
 RT filament gene family."
 RL EMBO J. 6:1617-1626(1987).
 RN [2]
 RP SEQUENCE OF ONE OF THE 13 RESIDUE REPEATS.

RA MEDLINE=88158120; PubMed=2450354;
 RA Lee V.M.-Y., Otvos L. Jr., Carden M.J., Hollosi M., Dietzschold B.,
 RA Lazzarini R.A.;
 RT "Identification of the major multiphosphorylation site in mammalian
 RT neurofilaments."
 RL Proc. Natl. Acad. Sci. U.S.A. 85:1998-2002(1988).
 CC -1- FUNCTION: NEUROFILAMENTS USUALLY CONTAIN THREE IF PROTEINS: L, M,
 CC AND H WHICH ARE INVOLVED IN THE MAINTENANCE OF NEURONAL CALIBER.
 CC -1- PTM: THERE ARE A NUMBER OF REPEATS OF THE TRIPLETTIDE K-S-P, NFM IS
 CC PHOSPHORYLATED ON A NUMBER OF THE SERINES IN THIS MOTIF. IT IS
 CC THOUGHT THAT PHOSPHORYLATION OF NFM RESULTS IN THE FORMATION OF
 CC INTERFILAMENT CROSS BRIDGES THAT ARE IMPORTANT IN THE MAINTENANCE
 CC OF AXONAL CALIBER.
 CC -1- PTM: PHOSPHORYLATION SEEMS TO PLAY A MAJOR ROLE IN THE FUNCTIONING
 CC OF THE LARGER NEUROFILAMENT POLYPEPTIDES (NF-M AND NF-H), THE
 CC LEVELS OF PHOSPHORYLATION BEING ALTERED DEVELOPMENTALLY AND
 CC COINCIDENT WITH A CHANGE IN THE NEUROFILAMENT FUNCTION.
 CC -1- SIMILARITY: BELONGS TO THE INTERMEDIATE FILAMENT FAMILY.

```

CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: Y00067; CAA68276.1; -.
DR PIR: A27864; A27864.
DR PIR: A30157; A30157.
DR MIM: 162250; -.
DR INTERPRO: IPR001664; -.
DR PFAM: PF00038; filament. 1.
DR PROSITE: PS00226; IF: 1.
KW Intermediate filament; Heptad repeat pattern; Coiled coil; Neurone;
KW Phosphorylation; Glycoprotein.
FT INI_MET 0
FT DOMAIN 1 103 HEAD.
FT DOMAIN 104 411 ROD.
FT DOMAIN 412 915 TAIL.
FT DOMAIN 104 135 COIL_1A.
FT DOMAIN 136 148 LINKER 1.
FT DOMAIN 149 247 COIL_1B.
FT DOMAIN 248 264 LINKER 12.
FT DOMAIN 265 286 COIL_2A.
FT DOMAIN 287 290 LINKER 2.
FT DOMAIN 291 411 COIL_2B.
FT DOMAIN 613 690 6 X 13 AA TANDEM REPEATS.
FT CARBOHYD 46 46 O-LINKED (GLCNAC) (BY SIMILARITY).
FT CARBOHYD 430 430 O-LINKED (GLCNAC) (BY SIMILARITY).
SQ SEQUENCE 915 AA; 102317 MW; 5F5D3DE34C9D9E50 CRC64;

Query Match 8.4%; Score 90; DB 1; Length 915;
Best Local Similarity 23.5%; Pred. No. 14;
Matches 39; Conservative 26; Mismatches 67; Indels 34; Gaps 6;

QY 26 KPIIDNYLHDKDKDEKIEQYDKNVKQASDKKQAKPOIPKSKSVAGYIEIP----- 79
DB 701 KEEEEKVEEVAPEKEVEK-----KEEKPKDVEKKKAESPVEEAEEVTTTKSVKVN 755
QY 80 -DADIKPEVYPGPATPEQLNRGVFAEENESLD-----DONISIAHGTFIDRPNYQFT 131
DB 756 LEKETEKEEGKPLQOQKEKEKAGEGSEEGSKAKGSKKEIDIAVNGEV-----EGKEEV 811
QY 132 NTKAAKGSNVYKRVGNETRRKYMTSIRDKPPTDVGVLDEQKGKDK 177
DB 812 EDETKKESG-----GREEKGAVTNGLDLSPA-----DEKKGCGDK 846

RESULT 8
FKB3_YEAST STANDARD; PRT; 411 AA.
ID FKB3_YEAST
AC P38911;
DT 01-FEB-1995 (rel. 31, Created)
DT 01-NOV-1995 (rel. 32, Last sequence update)
DT 15-JUL-1999 (rel. 38, Last annotation update)
DE FK506-BINDING NUCLEAR PROTEIN (PEPTIDYL-PROLYL CIS-TRANS ISOMERASE)
DE (PPIASE) (EC 5.2.1.8) (PROLINE ROTAMASE) (NUCLEOLAR PROLINE
DE ISOMERASE) (FKBP-70).
GN FPR3 OR NP146 OR YML074C.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycetes; Saccharomycetales;
OC Saccharomycetaceae; Saccharomycetes.
RN (1)
RN SEQUENCE FROM N.A.
RC STRAIN=YN 214;
RX MEDLINE=95050937; PubMed=7525596;
RA Benton B.M., Zeng J.-H., Thorner J.;
RT "A novel FK506- and rapamycin-binding protein (FPR3 gene product) in
the yeast Saccharomyces cerevisiae is a proline rotamase localized to

```

```

RT the nucleolus.";
RL J. Cell Biol. 127:623-639(1994).
RN (2)
RN SEQUENCE FROM N.A.
RP STRAIN=S288C;
RX MEDLINE=94327703; PubMed=8051210;
RA Shan X., Xue Z., Melese T.;
RT "yeast NP146 encodes a novel prolyl cis-trans isomerase that is
RT located in the nucleolus.";
RL J. Cell Biol. 126:853-862(1994).
RN (3)
RN SEQUENCE FROM N.A.
RX MEDLINE=95010686; PubMed=7925954;
RA Manning-Krieg U.C., Henriquez R., Cammas F., Graff P.,
RA Gaveriaux S., Moya N.R.;
RT "Purification of FKBP-70, a novel immunophilin from Saccharomyces
RT cerevisiae, and cloning of its structural gene, FPR3.";
RL FEBS Lett. 352:98-103(1994).
RN (4)
RN SEQUENCE FROM N.A.
RC STRAIN=S288C / AB972;
RA Brown D., Bowman S., Barrell B.G., Rajandream M.A.;
RL Submitted (Oct-1994) to the EMBL/Genbank/DBD databases.
RN (5)
RN PHOSPHORYLATION OF TYR-184 AND SER-186.
RX MEDLINE=97294697; PubMed=9148902;
RA Wilson L.K., Dhillon N., Thorner J., Martin G.S.;
RT "Casein kinase II catalyzes tyrosine phosphorylation of the yeast
RT nucleolar immunophilin Fpr3.";
RL J. Biol. Chem. 272:12961-12967(1997).
CC -I- FUNCTION: PPIASES ACCELERATE THE FOLDING OF PROTEINS. FK506- AND
CC RAPAMYCIN-BINDING PROTEIN. SPECIFICALLY BINDS NUCLEAR LOCALIZATION
CC SEQUENCES. MAY BE INVOLVED IN THE ASSEMBLY OR FOLDING OF RIBOSOMAL
CC PROTEINS.
CC -I- CATALYTIC ACTIVITY: CIS-TRANS ISOMERIZATION OF PROLINE IMIDIC
CC PEPTIDE BONDS IN OLIGOPEPTIDES.
CC -I- ENZYME REGULATION: INHIBITED BY BOTH FK506 AND RAPAMYCIN.
CC -I- SUBCELLULAR LOCATION: NUCLEAR; NUCLEOLUS.
CC -I- PTM: PHOSPHORYLATED AT TYROSINE AND DEPHOSPHORYLATED BY THE
CC PHOSPHOTYROSINE-SPECIFIC PHOSPHOPROTEIN PHOSPHATASE PP1.
CC -I- SIMILARITY: BELONGS TO THE FKBP-TYPE PPIASE FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: L34569; AAB04165.1; -.
DR EMBL: X79379; CAA55924.1; -.
DR EMBL: S73876; AAB31995.1; -.
DR EMBL: Z46373; CAA86504.1; -.
DR PIR: S47927; S47927.
DR PIR: A53544; A53544.
DR HSSP: P20081; 1YAT.
DR SGD: S0004539; NP146.
DR INTERPRO: IPR001179; -.
DR PFAM: PF00254; FKBP; 1.
DR PROSITE: PS00453; FKBP_PPIASE_1; 1.
DR PROSITE: PS00454; FKBP_PPIASE_2; 1.
DR PROSITE: PS50059; FKBP_PPIASE_3; 1.
KW Isomerase; Rotamase; Nuclear protein; Phosphorylation.
FT DOMAIN 60 87 ASP/GLU-RICH (HIGHLY ACIDIC).
FT DOMAIN 88 99 LYS-RICH (HIGHLY BASIC).
FT DOMAIN 101 119 ASP/GLU-RICH (HIGHLY BASIC).
FT DOMAIN 173 248 ASP/GLU-RICH (HIGHLY ACIDIC).
FT DOMAIN 250 298 LYS-RICH (HIGHLY BASIC).
FT DOMAIN 256 271 NUCLEAR LOCALIZATION SIGNAL (POTENTIAL).
FT DOMAIN 324 411 PPIASE, FKBP-TYPE.
FT MOD_RES 184 184 PHOSPHORYLATION (BY CK2).
FT MOD_RES 186 186 PHOSPHORYLATION (BY CK2).

```



```

OY 54 -SKDK-----KQAKPQIP-----KDKSVAGYIEIPDAIKEP 86
DB 1900 VOLDDIVSYLNHTNATNPILPILFNYDNEVYVLLPMPSTPKRYVAFYIKNP----- 1952
OY 87 VYPGATPPEQLNRGVSAFENESLDDONISIAGHFFIDRPNYQFTNLKAAK 146
DB 1953 -THFVQVEEATNLLISFLHYPKTTDLNOLPPNNTSESTRPGQTSFTLTNKLSEPKR- 2010
OY 147 GNETRKYKMTSIRDVCPFDVGLDEQKQKQKQTLITCDYN 188
DB 2011 -----KPAVTGLM-----PKSQSIITLSDTJN 2031

RESULT 13
TOP1_DROME STANDARD: PRT; 972 AA.
AC P30189;
DT 01-APR-1993 (Rel. 25, Created)
DT 01-APR-1993 (Rel. 25, Last sequence update)
DT 01-OCT-2000 (Rel. 40, Last annotation update)
DE DNA TOPOISOMERASE I (EC 5.99.1.2).
GN TOP1.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=93117086; PubMed=1335568;
RA Hsieh T.-S., Brown S.D., Huang P., Postel J.;
RT "Isolation and characterization of a gene encoding DNA topoisomerase
RT I in Drosophila melanogaster."
RL Nucleic Acids Res. 20:6177-6182(1992).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=OREGON-R;
RX MEDLINE=96354910; PubMed=8769417;
RA Zhang C.X., Lee M.P., Chen A.D., Brown S.D., Hsieh T.-S.;
RT "Isolation and characterization of a Drosophila gene essential for
RT early embryonic development and formation of cortical cleavage
RT furrows."
RL J. Cell Biol. 134:923-934(1996).
CC -I- FUNCTION: THE REACTION CATALYZED BY TOPOISOMERASES LEADS TO THE
CC CONVERSION OF ONE TOPOLOGICAL ISOMER OF DNA TO ANOTHER.
CC -I- CATALYTIC ACTIVITY: ATP-INDEPENDENT BREAKAGE OF SINGLE-STRANDED
CC DNA, FOLLOWED BY PASSAGE AND REJOINING.
CC -I- MISCELLANEOUS: EUKARYOTIC TOPOISOMERASE I AND II CAN RELAX BOTH
CC NEGATIVE AND POSITIVE SUPERCOILS, WHEREAS PROKARYOTIC ENZYMES
CC RELAX ONLY NEGATIVE SUPERCOILS.
CC -I- MISCELLANEOUS: WHEN A TOPOISOMERASE TRANSIENTLY BREAKS A DNA
CC BACKBONE BOND, IT SIMULTANEOUSLY FORMS A PROTEIN-DNA LINK, IN
CC WHICH A TYROSYL OXYGEN IN THE ENZYME IS JOINED TO A DNA PHOSPHORUS
CC AT ONE END OF THE ENZYME-SEVERED DNA STRAND.
CC -I- SIMILARITY: BELONGS TO THE EUKARYOTIC TYPE I TOPOISOMERASE FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
DR EMBL: M74557; AAA28951.1; -.
DR EMBL: U80064; AAC24158.1; -.
DR PIR: S35521; S35521.
DR HSSP: P11387; 1A35.
DR FLYBASE: FBgn0004924; Top1.
DR INTERPRO: IPR001631; -.
DR PFAM: PF01028; Topoisomerase_I; 1.
DR PRINTS: PR00416; EUPPISMRASEI
DR PROSITE: PS00176; TOPOISOMERASE_I_EUK; 1.
KW Isomerase; Topoisomerase; DNA-binding.

```

```

FT DOMAIN 32 39 POLY-HIS.
FT DOMAIN 40 198 SER-RICH.
FT ACT_SITE 930 930 DNA_CLEAVAGE (BY SIMILARITY).
SQ SEQUENCE 972 AA; 11688 MW; 3764B8BDEEFA30CD CRC64;

Query Match
Best Local Similarity 7.9%; Score 85; DB 1; Length 972;
Matches 41; Conservative 28; Mismatches 86; Indels 26; Gaps 6;

OY 30 DNYLHDKDKDEIEGYDYNVKEQASKDKKQ-----QAKPQIPKSKSVAGYIEIPDAIK 84
DB 140 DKERRDKDKRGSSSSSRHKSSSRDKERSSSHKSSSSSSSKSRHSSSSSSSK 199
OY 85 E-PVY-----PGPATPEQLNRGVSAFENESLDDONISIAGHFFIDRPNYQFTNLKAAK 137
DB 200 DHPSYDGVFAKPEPYSQLHMSGVDATQOMQLSGYEAAAGTNTNGCNVAGANYKNGY 259
OY 138 KGSWYFVYGVNETRK-----YKMTSIRDVCPFDVGLDEQKQKQKQTLITCDYN 188
DB 260 ERSIVDIKKKEESFNLLSQASSCDYSMSQFPADEPPFV-VKHEQSYAREDSSTM-----NYN 314
OY 189 E 189
DB 315 D 315

RESULT 14
YACE_BRELA
ID YACE_BRELA STANDARD: PRT; 195 AA.
AC P56187;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE HYPOTHETICAL PROTEIN IN PFG 3'REGION (FRAGMENT).
OS Brevibacterium lactofermentum.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Corynebacterineae.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 13869;
RA Yoon K.H.;
RL Submitted (NOV-1993) to the EMBL/GenBank/DBJ databases.
RN [2]
RP IDENTIFICATION.
RA Rudd K.E.;
RL Unpublished observations (JAN-1997).
CC -I- SIMILARITY: BELONGS TO THE UPF0038 FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
DR EMBL: L18875; -. NOT_ANNOTATED_CDS.
DR HSSP: P06143; 1HEX.
DR INTERPRO: IPR001977; -.
DR PFAM: PF01121; UPF0038; 1.
DR PROSITE: PS01294; UPF0038; 1.
KW Hypothetical protein; ATP-binding.
FT NP_BIND 8 15 ATP (POTENTIAL).
FT TER 195 195
SQ SEQUENCE 195 AA; 21135 MW; D7F0333860E0BDD3 CRC64;

Query Match
Best Local Similarity 7.9%; Score 84.5; DB 1; Length 195;
Matches 44; Conservative 21; Mismatches 55; Indels 35; Gaps 10;

OY 65 IPDKSKVAGYIE-----IPDAD--IKEPVYPG-PATPEQLNR-GVSAFENESLDDONI 115

```

Db 10 ISSGKSTVALLSSEGLFLIIDADQIAEIVEPQALAEVAFGEPIIKEDSLDQ-- 67
 QY 116 SIAGHTFIDRPNYQFTNLKAAGKSMYFKVGNET--RKYKMTSIRDYKPTDVGILD---- 170
 Db 68 GLAKAFVDEHETALLN-----SIHPRIAETARF---AEANGTKVAIYDMLPL 117
 QY 171 EOKGDKQLTLITCDYN-----EKTGVMEK 196
 Db 118 VDKGLDRGMDLVLVVDYVNEERVRLVEKRGLEK 152

RESULT 15
 FKBA_DROME STANDARD; PRT; 357 AA.

ID FKBA_DROME STANDARD; PRT; 357 AA.
 AC P54397;
 DT 01-OCT-1996 (Rel. 34, Created)
 DT 01-OCT-1996 (Rel. 34, Last sequence update)
 DT 15-JUL-1999 (Rel. 38, Last annotation update)
 DE 39 KDA FK506-BINDING NUCLEAR PROTEIN (PEPTIDYL-PROLYL CIS-TRANS ISOMERASE) (PIPIASE) (EC 5.2.1.8).
 GN FK506-BP1 OR FKBP39.
 OS Drosophila melanogaster (Fruit fly).
 OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-CANTON-S;
 RX MEDLINE=95278752; PubMed=7538962;
 RA Theopold U., Dal Zotto L., Hultmark D.;
 RT "FKBP39, a Drosophila member of a family of proteins that bind the immunosuppressive drug FK506.";
 RL Gene 156:247-251(1995).
 CC -!- FUNCTION: PIPIASES ACCELERATE THE FOLDING OF PROTEINS. FKBP506 MAY FUNCTION IN A SIGNAL TRANSDUCTION CASCADE DURING EARLY DEVELOPMENT.
 CC -!- CATALYTIC ACTIVITY: CIS-TRANS ISOMERIZATION OF PROLINE IMIDIC PEPTIDE BONDS IN OLIGOPEPTIDES.
 CC -!- SUBCELLULAR LOCATION: NUCLEAR (PROBABLE).
 CC -!- TISSUE SPECIFICITY: UBIQUITOUSLY EXPRESSED, HIGHEST LEVELS IN OVARY.
 CC -!- DEVELOPMENTAL STAGE: EXPRESSED DURING ALL STAGES OF DEVELOPMENT WITH HIGHEST EXPRESSION IN EARLY EMBRYO.
 CC -!- SIMILARITY: BELONGS TO THE FKBP-TYPE PIPIASE FAMILY.
 CC -!- CAUTION: IT IS UNCERTAIN WHETHER MET-1 OR MET-3 IS THE INITIATOR.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announce/ or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: Z46894; CA86996.1; -.
 DR HSSP: P27124; IROU.
 DR FLXBASE: FBgn0013269; FK506-bp1.
 DR INTERPRO: IPR001179; -.
 DR PRAM: PF00254; FKBP_1.
 DR PROSITE: PS00453; FKBP_PPIASE_1; FALSE_NEG.
 DR PROSITE: PS00454; FKBP_PPIASE_2; 1.
 DR PROSITE: PS00059; FKBP_PPIASE_3; 1.
 KM Isomerase; Rotamase; Nuclear protein.
 FT DOMAIN 89 99 ASP/GLU-RICH (HIGHLY ACIDIC).
 FT DOMAIN 119 183 ASP/GLU-RICH (HIGHLY ACIDIC).
 FT DOMAIN 186 247 LYS-RICH (BASIC).
 FT DOMAIN 269 357 PIPIASE, FKBP-TYPE.
 SQ SEQUENCE 357 AA: 39428 MW: F12C6431D4C3B55F CRC64;

Query Match 7.9%; Score 84.5; DB 1; Length 357;
 Best Local Similarity 27.6%; Pred. No. 12;

Matches 40; Conservative 18; Mismatches 62; Indels 25; Gaps 7;
 QY 47 KNVEQASDKDKQOAKPOIPKDSKVAGYIEIPDADIKEPYPGPAT--PEQLNRGVSA 104
 Db 200 KSGKEQNGVAKKEEAKQO-QKKKEKPEAKKEQPKA--KEPAKQOPASKDPRITIGGKIY 256
 QY 105 EE-----NESLDQONISAGHTFIDRPNYQFTNLKAAGKSMYFKVGNETRYKMTSIR 159
 Db 257 DQVVGKGEAEAKQKGRVSV---YYIGRLQSNKKTFTDSLTKGKPPKFAFGG-----E 304
 QY 160 DVKPTDYGVLDEQKGRDKQLTLITC 184
 Db 305 VIRGMDVGVAGMKVGGRR---VITC 326

Search completed: March 2, 2001, 10:24:33
 Job time: 120 sec

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 2, 2001, 10:20:53 ; Search time 15.51 seconds
(without alignments)
901.840 Million cell updates/sec

Title: US-09-292-437-3
Perfect score: 1076
Sequence: 1 MKKWNRLMTIAGVVLIVA.....YNEKGVWEKKRIFVATEVK 206

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 195891 seqs, 67900655 residues

Total number of hits satisfying chosen parameters: 195891

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :
1: PIR_66:*
2: PIR1:*
3: PIR3:*
4: PIR4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	104	9.7	352	2 T36719	probable integral
2	102	9.5	552	2 T25593	hypothetical prote
3	101.5	9.4	1089	2 T31583	hypothetical prote
4	97.5	9.1	678	2 A54514	glutamic acid-rich
5	97	9.0	887	2 S59777	hypothetical prote
6	95.5	8.9	644	2 S55395	neurofilament prot
7	94.5	8.8	993	2 S48436	probable RNA helic
8	93.5	8.7	763	2 T08929	hypothetical prote
9	93	8.6	253	2 S73653	hypothetical prote
10	91.5	8.5	312	2 T25894	hypothetical prote
11	91.5	8.5	485	2 S73333	hypothetical prote
12	91.5	8.5	845	2 A45669	Mg260 homolog - My
13	91.5	8.5	1288	2 T46486	neurofilament trip
14	91.5	8.5	198	2 T669823	chromosomal protei
15	91	8.5	1290	2 A55094	hypothetical prote
16	90	8.4	916	2 A27864	chromosomal protei
17	90	8.4	1192	2 A71623	neurofilament trip
18	89.5	8.3	411	2 S48647	probable secreted
19	89.5	8.3	1609	2 H72013	peptidylprolyl iso
20	89	8.3	359	2 H64547	polymorphic membra
21	89	8.3	359	2 E71960	peptide methionine
22	89	8.3	472	2 T27903	probable peptide m
23	89	8.3	641	2 S41861	hypothetical prote
24	89	8.3	691	2 T32748	gene NS-1 protein
25	89	8.3	1388	2 T17269	hypothetical prote
26	89	8.3	3488	2 T34418	hypothetical prote
27	88.5	8.2	1516	2 E71619	hypothetical prote
28	88	8.2	584	2 A69980	RAD2 endonuclease
29	88	8.2	849	2 S00030	penicillin-binding
					neurofilament trip

30	88	8.2	1291	2 T22382	hypothetical prote
31	87.5	8.1	464	2 T35943	probable hydrolyti
32	87	8.1	605	2 T02350	hypothetical prote
33	87	8.1	712	1 S69782	outer membrane pro
34	87	8.1	1151	2 T24541	hypothetical prote
35	87	8.1	1909	2 A45592	liver stage antige
36	87	8.1	2078	2 T09326	hypothetical prote
37	86.5	8.0	888	2 T25713	legument protein -
38	85	7.9	507	2 C82901	conserved hypotet
39	85	7.9	608	2 T22897	hypothetical prote
40	85	7.9	627	2 T04562	hypothetical prote
41	85	7.9	665	2 B71609	hypothetical prote
42	85	7.9	972	2 S35521	DNA topoisomerase
43	84.5	7.9	253	2 T36718	probable integral
44	84.5	7.9	357	2 J04090	FK506-binding 39k
45	84.5	7.9	641	2 S41439	gene NS-1 protein

ALIGNMENTS

RESULT 1
T36719
probable integral membrane protein - Streptomyces coelicolor
C:Species: Streptomyces coelicolor
C>Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 03-Dec-1999
C:Accession: T36719
R:Murphy, L.; Harris, D.; Bentley, S.D.; Parkhill, J.; Barrell, B.G.; Rajandream, M.A
submitted to the EMBL Data Library, June 1999
A:Reference number: Z21612
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-352 <MUR>
A:Cross-references: EMBL:AL079308; PDB:CA045217.1; GSPDB:GN00070; SCODEB:SCH69.20C
A:Experimental source: strain A3(2)
C:Genetics:
A:Gene: SCODEB:SCH69.20C

Query Match 9.7%; Score 104; DB 2; Length 352;
Best Local Similarity 23.2%; Pred. No. 0.61;
Matches 47; Conservative 40; Mismatches 82; Indels 34; Gaps 11;

OY 13 GVVLLIIVAALFAKPHNDYLNHDKDEKTEOYDKNVKEQASDKKQQAQPIPKKSKV 72
DB 148 GVVLLLEVTYQILMWTNVRH-----AQNQAASNDQDWANGKRS-----PGSFEFG 194

OY 73 AGY--IEIPDADIKEPYPGAPATPEQLNRGV--SFAEE--NESLDDQ---NISAGHTEI 123
DB 195 QGFALHIFKLDVVVFIAEISSKVLDRGMVGHVXEDGLKTAMPDAKAGNFGLAGHRT 254

OY 124 DRPNYQFTNLKAAKKSMYFKVGNETRKTKMYSIRDV-KPTDGVGLDEO-----KGKD 176
DB 255 HGEPEFRYN--KLEPGDPVIVETQDKYFVKMASILPVSPSVSLDVPVKOSGFGKPG 312

OY 177 KQLTLITC-DDYNEKTS--VWEK 196
DB 313 RYTLTCTCTPEFTSKRYMIVWGK 335

RESULT 2
T25593
hypothetical protein C32E12.1 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C>Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999
C:Accession: T25593
R:Wlcox, L.
submitted to the EMBL Data Library, November 1996
A:Description: The sequence of C. elegans cosmid C32E12.
A:Reference number: Z20055
A:Accession: T25593
A:Status: preliminary; translated from GB/EMBL/DBJ

QY 78 -----IPDADIKPEVYPGPATPEQJLNRGVSEAEENESLDD-ONISIAGHTFIDRPN 127
Db 380 NREKFTQJLDQDVDTREKJLHSHSKSVKKLQK--QLOKDEKVDLKNVPANSQKIIAEET 437
QY 128 YQFTNLKAAKKGSMVYFKVGNETRKYKMTSIRD-VKPTDVGVLDEQKQKQJTLIT 183
Db 438 NKKDLEKQK-----EKREKJLKNVMDSLKKTQJLOEKREKVEKEKELMEIS 483

Search completed: March 2, 2001, 10:22:50
Job time: 117 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 2, 2001, 10:22:03 ; Search time 23.18 Seconds

(without alignments)
1041.623 Million cell updates/sec

Title: US-09-292-437-3

Perfect score: 1076

Sequence: 1 MKKWTNRLMTIAGVILVA.....YNEKTVGEKRIKIFVATEVK 206

Scoring table: BLOSUM62

Searched: 374700 seqs, 117207915 residues

Total number of hits satisfying chosen parameters: 374700

Minimum DB seq length: 0
Maximum DB seq length: 200000000
Post-processing: Minimum Match 08
Maximum Match 1008
Listing first 45 summaries

- Database :
- 1: SP archaea:*
 - 2: SP bacteria:*
 - 3: SP fungi:*
 - 4: SP human:*
 - 5: SP invertebrate:*
 - 6: SP mammal:*
 - 7: SP mhc:*
 - 8: SP organelle:*
 - 9: SP phage:*
 - 10: SP plant:*
 - 11: SP rodent:*
 - 12: SP virus:*
 - 13: SP vertebrate:*
 - 14: SP unclassified:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the total score being printed, and is derived by analysis of the score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1076	100.0	206	2	09S446 staphylococ
2	184	17.1	193	2	09KBB3 bacillus ha
3	182.5	17.0	187	2	09KEV9 bacillus ha
4	133.5	12.4	194	2	09KBO9 bacillus ha
5	126	11.7	365	2	06B213 actinomyces
6	106.5	9.9	231	2	031340 bacillus ce
7	104	9.7	352	2	09XA14 streptomyce
8	102	9.5	552	5	P91122 caenorhabdi
9	101	9.4	1017	10	09SS568 arabidopsi
10	97.5	9.1	673	5	09UON1 plasmidum
11	97.5	9.1	810	6	077788 bos taurus
12	97	9.0	887	3	006106 saccharomye
13	96.5	9.0	1147	5	09Y1H3 dictyostell
14	94	8.7	643	5	09U229 caenorhabdi
15	93.5	8.7	763	10	09SUA1 arabidopsi
16	93	8.6	253	2	P75281 mycoplasma
17	92	8.6	2268	5	093515 caenorhabdi
18	91.5	8.5	312	5	P91570 mycoplasma
19	91.5	8.5	485	2	P75139 mycoplasma

20	91.5	8.5	845	11	063370 ratius norv
21	91.5	8.5	1202	4	09UNT9 homo sapien
22	91.5	8.5	1288	4	095752 homo sapien
23	91.5	8.5	1288	4	09NTJ3 homo sapien
24	91	8.5	954	10	P93826 arabidopsi
25	90.5	8.4	207	2	09K558 bacillus ha
26	90	8.4	916	4	09UE07 homo sapien
27	90	8.4	1192	5	09E127 plasmidum
28	89.5	8.3	1609	2	09RB58 chlamydia p
29	89.5	8.3	1609	2	09Z605 chlamydia p
30	89	8.3	348	10	039448 ciocer ariet
31	89	8.3	359	2	085224 helicobacte
32	89	8.3	359	2	09ZMK8 helicobacte
33	89	8.3	641	12	065023 aleutian mi
34	89	8.3	691	5	044741 caenorhabdi
35	89	8.3	978	4	075032 homo sapien
36	89	8.3	1388	4	09UFR9 streptococ
37	89	8.3	3488	5	P91257 caenorhabdi
38	88.5	8.2	309	2	09R6P5 streptococ
39	88.5	8.2	309	2	P72538 streptococ
40	88.5	8.2	978	5	09U6W5 caenorhabdi
41	88.5	8.2	1516	5	096154 plasmidum
42	88	8.2	584	2	032032 bacillus su
43	88	8.2	1291	5	09XV10 caenorhabdi
44	87.5	8.1	241	2	09RPE0 mycoplasma
45	87.5	8.1	309	2	09L5X4 streptococ

ALIGNMENTS

RESULT	ID	Query Match	Score	DB 2	Length	206
1	09S446	100.0%	1076	2	206	
AC	09S446	Best local similarity	100.0%	Pred. No. 4.3e-76	Matches 206	Conservative 0; Mismatches 0; Indels 0; Gaps 0;
DT	01-MAY-2000 (TREMBLrel. 13, Created)					
DT	01-MAY-2000 (TREMBLrel. 13, Last sequence update)					
DT	01-MAY-2000 (TREMBLrel. 13, Last annotation update)					
DE	SORTASE.					
GN	SRFA.					
OS	Staphylococcus aureus.					
OC	Bacteria; Firmicutes; Bacillus/Clostridium group;					
OC	Bacillus/Staphylococcus group; Staphylococcus.					
OX	NCBI_Taxid=1280;					
RN	[1]					
RP	SEQUENCE FROM N.A.					
RC	STRAIN=8325-4;					
RX	MEDLINE=99357874; PubMed=10427003;					
RA	Mazmanian S.K., Liu G., Ton-That H., Schneewind O.;					
RT	"Staphylococcus aureus sortase, an enzyme that anchors surface					
RT	proteins to the cell wall.";					
RL	Science 285:760-763(1999).					
DR	EMBL; AF162687; AAD48437.1;					
SQ	SEQUENCE 206 AA; 23541 MW; DC3B65C51E145C7B CRC64;					

Query Match	Score	DB 2	Length	206
Best local similarity	100.0%	Pred. No. 4.3e-76	Matches 206	Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY	1	MKKWTNRLMTIAGVILVAAYLFAPPHIDNYLHDKDKDEKIEQYDKNVKEQSKDKQKQ 60		
DB	1	MKKWTNRLMTIAGVILVAAYLFAPPHIDNYLHDKDKDEKIEQYDKNVKEQSKDKQKQ 60		
QY	61	AKQPIRKSKSVAGYIEIPDADIKFVYVGPAPPEQLNKGVSFAEENESIDDOINISIAH 120		
DB	61	AKQPIRKSKSVAGYIEIPDADIKFVYVGPAPPEQLNKGVSFAEENESIDDOINISIAH 120		
QY	121	TFIDRPNOFTNLKAQKSMYFKGNETRKXKMSIRDVKPTDVGVLDEQKQKQKLT 180		
DB	121	TFIDRPNOFTNLKAQKSMYFKGNETRKXKMSIRDVKPTDVGVLDEQKQKQKLT 180		
QY	181	LITCDYNEKTVGEKRIKIFVATEVK 206		

OS *Caenorhabditis elegans*.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditiidae; Peloderinae; *Caenorhabditis*.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-BRISTOL N2;
RX MEDLINE=94150718; PubMed=7906398;
RA Wilson R., Ainscough R., Anderson K., Baynes C., Berks M.,
RA Bonfield J., Burton J., Connell M., Copest T., Cooper J., Coulson A.,
RA Craxton M., Dear S., Du Z., Durbin R., Favello A., Fulton L.,
RA Gardner A., Green P., Hawkins T., Hillier L., Jier M., Johnston L.,
RA Jones M., Kershaw J., Kirsten T., Lalister N., Latreille P.,
RA Lightning J., Lloyd C., McMurray A., Mortimore B., O'Callaghan M.,
RA Parsons J., Percy C., Rifken L., Roopra A., Saunders D., Showkeen R.,
RA Smaldon N., Smith A., Sonhammer E., Staden R., Sulston J.,
RA TERRY-MEG J., Thomas K., Vaudin M., Vaughan K., Waterston R.,
RA Watson A., Weinstock L., Wilkinson-Spratt J., Wohlman P.,
RT "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
RT *elegans*.";
RL Nature 368:32-38(1994).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-BRISTOL N2;
RA Wilcox L.;
RN Submitted (JAN-1997) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN-BRISTOL N2;
RA Waterston R.;
RN Submitted (NOV-1996) to the EMBL/GenBank/DBJ databases.
RN [4]
RP SEQUENCE FROM N.A.
RC STRAIN-BRISTOL N2;
RA Waterston R.;
RL Submitted (MAY-1997) to the EMBL/GenBank/DBJ databases.
DR EMBL; 080032; AB53878.1; -.
SQ RNA-directed DNA polymerase.
SEQUENCE 552 AA; 62427 MW; D570979E3BD7ED94 CRC64;

RESULT	9
Q9SS68	
ID	Q9SS68
AC	Q9SS68
DT	01-MAY-2000 (TREMBLrel. 13, Created)
DT	01-MAY-2000 (TREMBLrel. 13, Last sequence update)
DT	01-JUN-2000 (TREMBLrel. 14, Last annotation update)
DE	PUTATIVE PHOSPHORIBOSYLTRANSFERINATE TRANSFERASE.
GN	TLJ13.4.
OS	Arabidopsis thaliana (Mouse-ear cress).
OC	Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;

OC Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II;
 OC Brassicales; Brassicaceae; Arabidopsis.
 OX NCBI_TaxID=3702;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=CV, COLUMBIA;
 RA Lin X., Kaul S., Town C.D., Benito M., Creasy T.H., Haas B.,
 RA Romling C.M., Koo H., Fujii C.Y., Ultebrack T.R., Barnstead M.E.,
 RA Bowman C.L., White O., Nierman W.C., Fraser C.M.;
 RT "Arabidopsis thaliana chromosome II BAC T1213 genomic sequence."
 RL Submitted (Oct-1999) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AC009327; AF03465.1; -
 DR INTERPRO: IPR000008; -
 DR PFAM; PF00168; C2; 3.
 DR PROSITE; PS50004; C2_DOMAIN_2; 3.
 DR Transferrase.
 SQ SEQUENCE 1017 AA; 114265 MW; ECDB92DB6E279C94 CRC64;

	Query Match	9.4%;	Score 101;	DB 10;	Length 1017;
	Best Local Similarity	26.3%;	Pred. No. 9.1;		
	Matches	50;	Conservative	24;	Mismatches 72; Indels 44; Gaps 10;
OY	36	KDDKDEIEQYDNNVKEAKSKDKQOAKPOIPRD----	KSXVGYEIEIPDADIEQVYGP	91	
	177	LEKKKGDKKEEKKPREEAKPDEK----	KPDAPDPTAKKPDYNAVAPPPPAEYKKNPIIQK	233	
OY	92	A-----TPEDLNRCVSFAEENE--SLD-DONISAGHTFTIDRPNYQFTNLKAAK	138		
	234	AETVAKONELGIPPEVNNRODLIGSLEPLSLTRDQNRG--GGDLVDLRMPFLIRVAKAKR	292		
OY	139	-----GSMYEFKVGNETRKKYKMTSIRDVKPTDVGVLDEQKGD-KQITLITCDQVNE---	189		
	293	AKNDGSNPNYALALVIGTNGVKTRS-----	QTGKDMQDVAFAFEKESLNSTSL	338	
OY	190	KTGVEKKRTI	199		
	339	EVSVMSEERI	348		

RESULT	10	
Q90UN1		
ID	Q90UN1	PRELIMINARY;
AC	Q90UN1;	PRT; 673 AA.
DT	01-MAY-2000 (TREMBLrel. 13, Created)	
DT	01-MAY-2000 (TREMBLrel. 13, Last sequence update)	
DT	01-MAY-2000 (TREMBLrel. 13, Last annotation update)	
DE	HYPOTHETICAL GARP PROTEIN.	
CN	GARP.	
OS	Plasmodium falciparum.	
OC	Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.	
OX	NCBI_TaxID=5683;	
RN	[1]	
RP	SEQUENCE FROM N.A.	
RC	STRAIN=3D7;	
RA	Bowman S., Churcher C., Harris B., Harris D., Lawson D., Quail M.,	
RA	Summitted (SEP-1998) to the EMBL/GenBank/DBJ databases.	
RL	EMBL; AL031746; CAB63561.1; -.	
SR	SEQUENCE 673 AA; 79795 MW; 7A7FF6921D83DA26 CRC64;	

	9.1%,	Score 97.5;	DB 5;	Length 673;	
Query Match	Best Local Similarity	24.1%;	Pred.	No. 10;	
Matches	42;	Conservative	32;	Mismatches	73; Indels 27; Gaps 8;
OY	34	HDKDKEITETDYDKVWKEDQASRKKKQQAKPOLIPRKSVAGTIELPDADIKEPYPGPAT	93		
		: : : : : : : : :	:		
Dd	128	HKDDKEKRERKKDRKKRD-KKKKHKKKKHKKKD-----ENSEVMSLYTQTGHK	180		
OY	94	PEQLKRGVSFAEEN---ESLDDQNISIAHGHTFIDPNPQFTLKAAKGSMYLEVGNET	150		
		: : : : : : : :			
Dd	181	KK--NATEHGHEINLEYEMWSEINNNAQGGLLSLP-VGYREOGCGGILSSVN-PTTSNDT	235		

QY 172 QKGDQTLTTCDDYNEKTVGWEKK 198
 Db 378 KENIAKRFVEFYQDOYEKEKDEEETHK 404

RESULT 14

Q90229 PRELIMINARY; PRT; 643 AA.
 AC Q90229;
 DT 01-MAY-2000 (TREMblrel. 13, Created)
 DT 01-MAY-2000 (TREMblrel. 13, Last sequence update)
 DE Y56A3A.32 (TREMblrel. 13, Last annotation update)
 DE Y56A3A.32 PROTEIN.
 GN Y56A3A.32.
 OS Caenorhabditis elegans.
 OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
 OC Rhabditidae; Peioderinae; Caenorhabditis.
 OX NCBI_Taxid=6239;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Matthews L.;
 RL Submitted (JAN-1999) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP SEQUENCE FROM N.A.
 RA none;
 RT "genome sequence of the nematode C. elegans: A platform for
 investigating biology."
 RL Science 282:2012-2018(1998).
 DR EMBL: AL132860; CAB60511.1; -.
 DR INTERPRO: IPR001100; -.
 DR INTERPRO: IPR001327; -.
 DR PRAM: PF00070; PYR_redox; 1.
 DR PRINTS: PR00368; FADPNR.
 DR PRINTS: PR00411; PNDRTASEI.
 SQ SEQUENCE 643 AA; 71257 MW; EAC5A54980A5F75D CRC64;

Query Match 8.7%; Score 94; DB 5; Length 643;
 Best Local Similarity 28.1%; Pred. No. 18;
 Matches 36; Conservative 12; Mismatches 44; Indels 36; Gaps 6;

QY 34 HDRKDEKIEQYDKNVKEQAKKQAKPOIPKDKSVAGYIIPDADIEPYPGPA 92
 Db 113 HSKNHEKHQKHEE-KEHAPEKKEAPEKPEP-----AEPEPEPEPAQ 154
 QY 93 ----TPQQL--NRGVSAEENESLDDONISLIGH-----TFIDRPNYQFTWLKAAGSM 141
 Db 155 KQAEQPPQAEKQETKDAEPKEQYDDROTEAVAHARRAPAAAEPP-----APSTSKADA 208
 QY 142 YEKVGN 149
 Db 209 VEKIGEE 216

RESULT 15

Q9SU01 PRELIMINARY; PRT; 763 AA.
 AC Q9SU01;
 DT 01-MAY-2000 (TREMblrel. 13, Created)
 DT 01-MAY-2000 (TREMblrel. 13, Last sequence update)
 DT 01-OCT-2000 (TREMblrel. 15, Last annotation update)
 DE HYPOTHETICAL 85.2 KDA PROTEIN.
 GN T15N24.80 OR AT4G2630.
 OS Arabidopsis thaliana (Mouse-ear cress).
 OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
 OC Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II;
 OC Brassicales; Brassicaceae; Arabidopsis.
 OX NCBI_Taxid=3702;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Bevan M., Zimmermann W., Gruenisen A., Wambutt R., Bancroft I.,
 RA Mewes H.W., Mayer K.F.X., Lemcke K., Scheller C.;

RL Submitted (MAY-1999) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP SEQUENCE FROM N.A.
 RA EU Arabidopsis sequencing project;
 RL Submitted (AUG-1999) to the EMBL/GenBank/DBJ databases.
 RN [3]
 RP SEQUENCE FROM N.A.
 RA Zimmermann W., Gruenisen A., Wambutt R., Kalicki J., Wohlmann P.,
 RA Smith A., Mewes H.W., Lemcke K., Mayer K.F.X.;
 RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
 RN [4]
 RP SEQUENCE FROM N.A.
 RA EU Arabidopsis sequencing project;
 RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AL078465; CAB43859.1; -.
 DR EMBL: AL161565; CAB79518.1; -.
 KW Hypothetical protein
 SQ SEQUENCE 763 AA; 85249 MW; D67FAF5BD4DB480B CRC64;

Query Match 8.7%; Score 93.5; DB 10; Length 763;
 Best Local Similarity 24.5%; Pred. No. 24;
 Matches 48; Conservative 27; Mismatches 74; Indels 47; Gaps 8;

QY 25 AKPHIDNYLHD--KDKDEKIEQYDKNVKEQAKD---KKQAKPOIPKDKSVAGYIE 77
 Db 47 AEP--DNWEIDAQIKDKDEKAEDEKSEYVKKNEDNAETQKMEKVEYTKQGAQATNM 104
 QY 78 IPDADIEPYPGPATPEQLNRGVS-----FAENESLDDONISLIGHTFIDRPNYQFTN 132
 Db 105 DEDADGKK-----EQTDGVSVEDTYKNEVESKN-----NYAKDD 141
 QY 133 LKAAGKSMV--YFKVGNETRRKKMTSIRDVKPTDVG-----VLDEQKGDQTLTI 182
 Db 142 EKETKEDITADHKKAKGKEDIQHADKANCTKGNATGDIKEGTIVDEKGTMDKEVE 201
 QY 183 TCDDYNEKTVGWEKK 198
 Db 202 NGDENKQVNEVEKKEK 217

Search completed: March 2, 2001, 10:23:33
 Job time: 90 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 2, 2001, 10:22:13 ; Search time 43.08 Seconds

(without alignments)
734.072 Million cell updates/sec

Title: US-09-292-437-3

Perfect score: 1076
Sequence: 1 MKKWTNRLMTIAGVILVA.....YNEKGVWEKRIIVATEVK 206

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 957798 seqs, 153513742 residues

Total number of hits satisfying chosen parameters: 957798

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Pending_Patents_AA:*
1: /cgn2_6/ptodata/2/paa/PCTUS.COMB.pep:*
2: /cgn2_6/ptodata/2/paa/US06.COMB.pep:*
3: /cgn2_6/ptodata/2/paa/US07.COMB.pep:*
4: /cgn2_6/ptodata/2/paa/US08.COMB.pep:*
5: /cgn2_6/ptodata/2/paa/US081.COMB.pep:*
6: /cgn2_6/ptodata/2/paa/US082.COMB.pep:*
7: /cgn2_6/ptodata/2/paa/US083.COMB.pep:*
8: /cgn2_6/ptodata/2/paa/US084.COMB.pep:*
9: /cgn2_6/ptodata/2/paa/US085.COMB.pep:*
10: /cgn2_6/ptodata/2/paa/US086.COMB.pep:*
11: /cgn2_6/ptodata/2/paa/US087.COMB.pep:*
12: /cgn2_6/ptodata/2/paa/US088.COMB.pep:*
13: /cgn2_6/ptodata/2/paa/US089.COMB.pep:*
14: /cgn2_6/ptodata/2/paa/US090.COMB.pep:*
15: /cgn2_6/ptodata/2/paa/US091.COMB.pep:*
16: /cgn2_6/ptodata/2/paa/US092.COMB.pep:*
17: /cgn2_6/ptodata/2/paa/US093.COMB.pep:*
18: /cgn2_6/ptodata/2/paa/US094.COMB.pep:*
19: /cgn2_6/ptodata/2/paa/US095.COMB.pep:*
20: /cgn2_6/ptodata/2/paa/US096.COMB.pep:*
21: /cgn2_6/ptodata/2/paa/US097.COMB.pep:*
22: /cgn2_6/ptodata/2/paa/US098.COMB.pep:*
23: /cgn2_6/ptodata/2/paa/US099.COMB.pep:*
24: /cgn2_6/ptodata/2/paa/US100.COMB.pep:*
25: /cgn2_6/ptodata/2/paa/US101.COMB.pep:*
26: /cgn2_6/ptodata/2/paa/US102.COMB.pep:*
27: /cgn2_6/ptodata/2/paa/US103.COMB.pep:*
28: /cgn2_6/ptodata/2/paa/US104.COMB.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1076	100.0	206	US-09-292-437-3	Sequence 3, Appli
2	1063	98.8	207	US-08-827-356-5465	Sequence 5465, Ap
3	1063	98.8	207	US-09-611-529-6656	Sequence 6656, Ap
4	1062	98.7	206	US-60-242-578-1023	Sequence 1023, Ap
5	1062	98.7	206	US-60-257-931-3336	Sequence 3336, Ap

6	794.5	73.8	203	18	US-09-450-969-4904	Sequence 4904, Ap
7	160.5	14.9	251	14	US-09-071-035-486	Sequence 486, Ap
8	160.5	14.9	251	15	US-09-134-000-5985	Sequence 5985, Ap
9	158.5	14.7	284	14	US-09-071-035-298	Sequence 298, Ap
10	158.5	14.7	284	20	US-09-634-341-29	Sequence 29, Appl
11	158.5	14.7	284	22	US-60-223-804-29	Sequence 29, Appl
12	158.5	14.7	284	27	US-09-292-437-6	Sequence 6, Appl1
13	158.5	14.7	315	15	US-09-134-000-6125	Sequence 6125, Ap
14	157.5	14.6	229	27	US-09-107-532-5472	Sequence 5472, Ap
15	156	14.5	227	27	US-09-292-437-4	Sequence 4, Appl1
16	154.5	14.4	197	14	US-09-071-035-488	Sequence 488, App
17	151.5	14.1	325	15	US-09-107-552-5753	Sequence 5753, Ap
18	150	13.9	150	20	US-09-634-341-37	Sequence 37, Appl
19	150	13.9	150	22	US-60-223-804-37	Sequence 37, Appl
20	146	13.6	256	14	US-09-071-035-300	Sequence 300, App
21	143	13.3	283	27	US-09-292-437-34	Sequence 34, Appl
22	140.5	13.1	246	27	US-09-292-437-7	Sequence 7, Appl
23	140	13.0	297	20	US-09-634-341-23	Sequence 23, Appl
24	140	13.0	297	22	US-60-223-804-23	Sequence 23, Appl
25	139.5	13.0	247	19	US-09-583-110-4595	Sequence 4595, Ap
26	139.5	13.0	249	27	US-09-107-433-2710	Sequence 2710, Ap
27	135.5	12.6	296	27	US-09-292-437-35	Sequence 35, Appl
28	129	12.0	248	12	US-08-827-356-3054	Sequence 3054, Ap
29	129	12.0	248	20	US-09-611-529-5479	Sequence 5479, Ap
30	126	11.7	283	20	US-09-634-341-8	Sequence 8, Appl1
31	126	11.7	283	22	US-60-223-804-8	Sequence 8, Appl1
32	126	11.7	365	27	US-09-292-437-5	Sequence 5, Appl1
33	125.5	11.7	304	22	US-09-634-341-26	Sequence 26, Appl
34	125.5	11.7	304	22	US-60-223-804-26	Sequence 26, Appl
35	125.5	11.7	304	27	US-09-292-437-36	Sequence 36, Appl
36	125	11.6	348	22	US-09-634-341-32	Sequence 32, Appl
37	125	11.6	348	22	US-60-223-804-32	Sequence 32, Appl
38	123	11.4	251	18	US-60-257-931-2943	Sequence 2943, Ap
39	116	10.8	243	15	US-09-134-000-6184	Sequence 6184, Ap
40	116	10.8	398	15	US-09-107-552-4688	Sequence 4688, Ap
41	115	10.7	236	15	US-09-107-532-6923	Sequence 6923, Ap
42	109.5	10.2	317	15	US-09-107-532-6714	Sequence 6714, Ap
43	108.5	10.1	305	20	US-09-634-341-6	Sequence 6, Appl1
44	108.5	10.1	305	22	US-60-223-804-6	Sequence 268, App
45	96	8.9	240	20	US-09-634-228-268	

ALIGNMENTS

RESULT 1
US-09-292-437-3
; Sequence 3, Application US/09292437
; GENERAL INFORMATION:
; APPLICANT: Olaf Schneewind
; APPLICANT: Sarkis Mazmanian
; APPLICANT: Gwen Liu
; APPLICANT: Hung Ton-That
; TITLE OF INVENTION: IDENTIFICATION OF SORTASE GENE
; FILE REFERENCE: 510015.213
; CURRENT APPLICATION NUMBER: US/09/292,437
; CURRENT FILING DATE: 1999-04-15
; NUMBER OF SEQ ID NOS: 36
; SOFTWARE: FASTSEQ For Windows Version 3.0
; SEQ ID NO 3
; LENGTH: 206
; TYPE: PRT
; ORGANISM: Staphylococcus aureus
US-09-292-437-3

Query Match 100.0%; Score 1076; DB 27; Length 206;
Best Local Similarity 100.0%; Pred. No. 3; 1e-96;
Matches 206; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 1 MKKWTNRLMTIAGVILVAAYLFAPKPHIDNYLHDKDKDKIIOYKKNKEQSKDKKQ 60
DB 1 MKKWTNRLMTIAGVILVAAYLFAPKPHIDNYLHDKDKDKIIOYKKNKEQSKDKKQ 60

US-09-611-529-6656

Query Match 98.8%; Score 1063; DB 20; Length 207;
Best Local Similarity 99.0%; Pred. No. 5.7e-95;
Matches 204; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 MKKWTNRLMTIAGVLLILVAAYLFAKPHIDNYLHDKDKDEKIEQYDKNKEQASKDKKQ 60
DB 2 MKKWTNRLMTIAGVLLILVAAYLFAKPHIDNYLHDKDKDEKIEQYDKNKEQASKDKKQ 61
QY 61 AKQIIPKDKSKVAGYIEIPDADIKEPVYPGPATPEQLNKGVSFAEENESLDDQNISIAGH 120
DB 62 AKQIIPKDKSKVAGYIEIPDADIKEPVYPGPATPEQLNKGVSFAEENESLDDQNISIAGH 121
QY 121 TPTDRPNYQFTNLKAAKKSVMYFKVGNETRKRYKMTSIRDVKTPTDVGVLDEQKDKKQ 180
DB 122 TPTDRPNYQFTNLKAAKKSVMYFKVGNETRKRYKMTSIRDVKTPTDVGVLDEQKDKKQ 181
QY 181 LITCDYNEKGTGWEKRIKIFVATEVK 206
DB 182 LITCDYNEKGTGWEKRIKIFVATEVK 207

RESULT 4

US-60-242-578-1023

; Sequence 1023, Application US/60242578

; GENERAL INFORMATION:

; APPLICANT: Haselbeck, R
; APPLICANT: Ohlsen, K. L.
; APPLICANT: Zyskind, J. W.
; TITLE OF INVENTION: Genes identified as essential in
; FILE REFERENCE: ELITRA.017PR2
; CURRENT APPLICATION NUMBER: US/60/242, 578
; NUMBER OF SEQ ID NOS: 1057
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1023
; LENGTH: 206
; TYPE: PRF
; ORGANISM: Staphylococcus aureus
US-60-242-578-1023

Query Match 98.7%; Score 1062; DB 22; Length 206;
Best Local Similarity 99.0%; Pred. No. 7.1e-95;
Matches 204; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 MKKWTNRLMTIAGVLLILVAAYLFAKPHIDNYLHDKDKDEKIEQYDKNKEQASKDKKQ 60
DB 1 MKKWTNRLMTIAGVLLILVAAYLFAKPHIDNYLHDKDKDEKIEQYDKNKEQASKDKKQ 60
QY 61 AKQIIPKDKSKVAGYIEIPDADIKEPVYPGPATPEQLNKGVSFAEENESLDDQNISIAGH 120
DB 61 AKQIIPKDKSKVAGYIEIPDADIKEPVYPGPATPEQLNKGVSFAEENESLDDQNISIAGH 120
QY 121 TPTDRPNYQFTNLKAAKKSVMYFKVGNETRKRYKMTSIRDVKTPTDVGVLDEQKDKKQ 180
DB 121 TPTDRPNYQFTNLKAAKKSVMYFKVGNETRKRYKMTSIRDVKTPTDVGVLDEQKDKKQ 180
QY 181 LITCDYNEKGTGWEKRIKIFVATEVK 206
DB 181 LITCDYNEKGTGWEKRIKIFVATEVK 206

RESULT 5

US-60-257-931-3336

; Sequence 3336, Application US/60257931

; GENERAL INFORMATION:

; APPLICANT: Haselbeck, R
; APPLICANT: Ohlsen, K. L.
; APPLICANT: Zyskind, J. W.

; APPLICANT: Trawick, JD

; APPLICANT: Wall, D

; TITLE OF INVENTION: Genes identified as essential in Staphylococcus aureus, Salmo

; TITLE OF INVENTION: typhimurium, Klebsiella pneumoniae and Pseudomonas aeruginos

; FILE REFERENCE: ELITRA.017PR4

; CURRENT APPLICATION NUMBER: US/60/257, 931

; NUMBER OF SEQ ID NOS: 3592

; SOFTWARE: FastSeq for Windows Version 4.0

; SEQ ID NO 3336

; LENGTH: 206

; TYPE: PRF

; ORGANISM: Staphylococcus aureus

US-60-257-931-3336

Query Match 98.7%; Score 1062; DB 28; Length 206;
Best Local Similarity 99.0%; Pred. No. 7.1e-95;
Matches 204; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 MKKWTNRLMTIAGVLLILVAAYLFAKPHIDNYLHDKDKDEKIEQYDKNKEQASKDKKQ 60
DB 1 MKKWTNRLMTIAGVLLILVAAYLFAKPHIDNYLHDKDKDEKIEQYDKNKEQASKDKKQ 60
QY 61 AKQIIPKDKSKVAGYIEIPDADIKEPVYPGPATPEQLNKGVSFAEENESLDDQNISIAGH 120
DB 61 AKQIIPKDKSKVAGYIEIPDADIKEPVYPGPATPEQLNKGVSFAEENESLDDQNISIAGH 120
QY 121 TPTDRPNYQFTNLKAAKKSVMYFKVGNETRKRYKMTSIRDVKTPTDVGVLDEQKDKKQ 180
DB 121 TPTDRPNYQFTNLKAAKKSVMYFKVGNETRKRYKMTSIRDVKTPTDVGVLDEQKDKKQ 180
QY 181 LITCDYNEKGTGWEKRIKIFVATEVK 206
DB 181 LITCDYNEKGTGWEKRIKIFVATEVK 206

RESULT 6

US-09-450-969-4904

; Sequence 4904, Application US/09450969

; GENERAL INFORMATION:

; APPLICANT: Lynn Doucette-Stamm et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCC
; FILE REFERENCE: PATH99-09A
; CURRENT APPLICATION NUMBER: US/09/450, 969
; NUMBER OF SEQ ID NOS: 7544
; SEQ ID NO 4904
; LENGTH: 203
; TYPE: PRF
; ORGANISM: S. epidermidis
US-09-450-969-4904

Query Match 73.8%; Score 794.5; DB 18; Length 203;
Best Local Similarity 72.2%; Pred. No. 6.1e-69;
Matches 148; Conservative 25; Mismatches 29; Indels 3; Gaps 1;

QY 1 MKKWTNRLMTIAGVLLILVAAYLFAKPHIDNYLHDKDKDEKIEQYDKNKEQASKDKKQ 60
DB 1 MKKWTNRLMTIAGVLLILVAAYLFAKPHIDNYLHDKDKDEKIEQYDKNKEQASKDKKQ 60
QY 61 AKQIIPKDKSKVAGYIEIPDADIKEPVYPGPATPEQLNKGVSFAEENESLDDQNISIAGH 120
DB 61 AKQIIPKDKSKVAGYIEIPDADIKEPVYPGPATPEQLNKGVSFAEENESLDDQNISIAGH 120
QY 121 TPTDRPNYQFTNLKAAKKSVMYFKVGNETRKRYKMTSIRDVKTPTDVGVLDEQKDKKQ 180
DB 121 TPTDRPNYQFTNLKAAKKSVMYFKVGNETRKRYKMTSIRDVKTPTDVGVLDEQKDKKQ 180
QY 181 LITCDYNEKGTGWEKRIKIFVATEVK 205
DB 181 LITCDYNEKGTGWEKRIKIFVATEVK 205

Db 178 LITCDVNEETGWTETKFIATQI 202

RESULT 7

US-09-071-035-486

Sequence 486, Application US/09071035

GENERAL INFORMATION:

APPLICANT: Gil H. Choi

TITLE OF INVENTION: Enterococcus faecalis Polynucleotides and Polypeptides

NUMBER OF SEQUENCES: 496

CORRESPONDENCE ADDRESS:

ADDRESSEE: Human Genome Sciences, Inc.

STREET: 9410 Key West Avenue

CITY: Rockville

STATE: Maryland

COUNTRY: USA

ZIP: 20850

COMPUTER READABLE FORM:

MEDIUM TYPE: Diskette, 3.50 inch, 1.4mb storage

COMPUTER: HP Vectra 486/33

OPERATING SYSTEM: MSDOS version 6.2

SOFTWARE: ASCII Text

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/071,035

FILING DATE:

CLASSIFICATION:

PRIOR APPLICATION DATA:

APPLICATION NUMBER:

ATTORNEY/AGENT INFORMATION:

NAME: A. Anders Brooks

REGISTRATION NUMBER: 36,373

REFERENCE/DOCKET NUMBER: PB369P2

TELECOMMUNICATION INFORMATION:

TELEPHONE: (301) 309-8512

TELEFAX: (301) 309-8504

INFORMATION FOR SEQ ID NO: 486:

SEQUENCE CHARACTERISTICS:

LENGTH: 251 amino acids

TYPE: amino acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: protein

US-09-071-035-486

Query Match 14.9%; Score 160.5; DB 14; Length 251;

Best Local Similarity 23.7%; Pred. No. 2.5e-07;

Matches 52; Conservative 46; Mismatches 104; Indels 17; Gaps 6;

QY 2 KKWNRMTIAGVYLIVAAVLFAPKPHIDNYLHDKD---EKIEQYD--KNVKEQASK 55

Db 14 KRGNWLNLSLVLEFIIGLALIFNNQIRSWVVOQNSRSYAVSKLPADVKKNMARETF 73

QY 56 D-----KKQAKPQIPKDKSKVAGYIEIPDADIKPPVPGPATPPQNLNGVSFAEEN 107

Db 74 DEDSVESLSTAVYKKAQFENKNLPYIGALIPSEINLPFKGLSNVALLT-GAGTMKED 132

QY 108 ESLDDNISAGHTFIDRPYQFTNLKAKKGSWYFVGNMTRKRYKMTSIRDVCPDVG 167

Db 133 QVMGKNNYALASHRTEDGVS-LFSPLETKRKDELITYTDLSTVYTKITSVEKIEPTVE 191

QY 168 VLDECKGDKQLTLITCDVNEKTGWMEKRKFVATEVK 206

Db 192 LIDVPGON-MITLTGCDLATRINAVGTLAATTPIK 229

RESULT 8

US-09-134-000-5985

Sequence 5985, Application US/09134000A

GENERAL INFORMATION:

APPLICANT: Lynn Doucette-Stamm et al

TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ENTEROCOCCUS

TITLE OF INVENTION: FAECALIS FOR DIAGNOSTICS AND THERAPEUTICS

FILE REFERENCE: GTC-005

CURRENT APPLICATION NUMBER: US/09/134,000A

CURRENT FILING DATE: 1998-08-13

NUMBER OF SEQ ID NOS: 6810

SEQ ID NO 5985

LENGTH: 251

ORGANISM: Enterococcus faecalis

US-09-134-000-5985

Query Match 14.9%; Score 160.5; DB 15; Length 251;

Best Local Similarity 23.7%; Pred. No. 2.5e-07;

Matches 52; Conservative 46; Mismatches 104; Indels 17; Gaps 6;

QY 2 KKWNRMTIAGVYLIVAAVLFAPKPHIDNYLHDKD---EKIEQYD--KNVKEQASK 55

Db 14 KRGNWLNLSLVLEFIIGLALIFNNQIRSWVVOQNSRSYAVSKLPADVKKNMARETF 73

QY 56 D-----KKQAKPQIPKDKSKVAGYIEIPDADIKPPVPGPATPPQNLNGVSFAEEN 107

Db 74 DEDSVESLSTAVYKKAQFENKNLPYIGALIPSEINLPFKGLSNVALLT-GAGTMKED 132

QY 108 ESLDDNISAGHTFIDRPYQFTNLKAKKGSWYFVGNMTRKRYKMTSIRDVCPDVG 167

Db 133 QVMGKNNYALASHRTEDGVS-LFSPLETKRKDELITYTDLSTVYTKITSVEKIEPTVE 191

QY 168 VLDECKGDKQLTLITCDVNEKTGWMEKRKFVATEVK 206

Db 192 LIDVPGON-MITLTGCDLATRINAVGTLAATTPIK 229

RESULT 9

US-09-071-035-298

Sequence 298, Application US/09071035

GENERAL INFORMATION:

APPLICANT: Gil H. Choi

TITLE OF INVENTION: Enterococcus faecalis Polynucleotides and Polypeptides

NUMBER OF SEQUENCES: 496

CORRESPONDENCE ADDRESS:

ADDRESSEE: Human Genome Sciences, Inc.

STREET: 9410 Key West Avenue

CITY: Rockville

STATE: Maryland

COUNTRY: USA

ZIP: 20850

COMPUTER READABLE FORM:

MEDIUM TYPE: Diskette, 3.50 inch, 1.4mb storage

COMPUTER: HP Vectra 486/33

OPERATING SYSTEM: MSDOS version 6.2

SOFTWARE: ASCII Text

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/071,035

FILING DATE:

CLASSIFICATION:

PRIOR APPLICATION DATA:

APPLICATION NUMBER:

ATTORNEY/AGENT INFORMATION:

NAME: A. Anders Brooks

REGISTRATION NUMBER: 36,373

REFERENCE/DOCKET NUMBER: PB369P2

TELECOMMUNICATION INFORMATION:

TELEPHONE: (301) 309-8512

TELEFAX: (301) 309-8504

INFORMATION FOR SEQ ID NO: 298:

SEQUENCE CHARACTERISTICS:

LENGTH: 284 amino acids

TYPE: amino acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: protein

US-09-071-035-298

Query Match

14.7%; Score 158.5; DB 14; Length 284;

Best Local Similarity 25.3%; Pred. No. 4.6e-07; Matches 56; Conservative 38; Mismatches 90; Indels 37; Gaps 7;

```
QY 1 MKKWTNRLMTIAGVLLIVA-----AYLFAKPHIDNYL-----HDKDKDE 40
  || | : : : : : || | : : : : :
Db 1 MMSKKRRRIIDGMILLIIGIGAFAYPFVSDALNNYLDQOIIAHYQAKASQENTKEMAE 60
  || | : : : : : || | : : : : :
QY 41 KIEQYDKNNKEQA-----SKDKQQAQAPQIPKDKSKVAGYIEIPDADIKPEVP 89
  || | : : : : : || | : : : : :
Db 61 LOEKMEKKNOELAKKGSNPGDLPFSETQKTKPKDKSYFESHIGVLTIPKINVRPIF- 119
  || | : : : : : || | : : : : :
QY 90 GPATPQOLNRGVSAFAENE---SLDDQNISIAHGTIFIDRPNYQFTMLKAAGKSMYFVK 146
  || | : : : : : || | : : : : :
Db 120 DKTNALLLEKGSLLGEGTSPGCTGTHAVISHRGLPQAKL-FTDLPKLKGDDEYIEY 178
  || | : : : : : || | : : : : :
QY 147 GNETRRYKMTSIRDVAPTDVGVLEQKGRDKQLTLITCDY 187
  || | : : : : : || | : : : : :
Db 179 NGKTLAYQVDQIKTVEPTDKDLHIESGD-LVTLLTCTPY 218
  || | : : : : : || | : : : : :
```

RESULT 10

US-09-634-341-29

; Sequence 29, Application US/09634341

; GENERAL INFORMATION:

; APPLICANT: Adderson, Elisabeth

; APPLICANT: Bohnsack, John

; TITLE OF INVENTION: GROUP B STREPTOCOCCUS POYPEPTIDES NUCLEIC ACIDS AND

; TITLE OF INVENTION: THERAPEUTIC COMPOSITIONS AND VACCINES THEREOF

; FILE REFERENCE: 2511-1-001

; CURRENT APPLICATION NUMBER: US/09/634,341

; CURRENT FILING DATE: 2000-08-08

; NUMBER OF SEQ ID NOS: 37

; SOFTWARE: Patent In Ver. 2.0

; SEQ ID NO 29

; LENGTH: 284

; TYPE: PRF

; ORGANISM: Enterococcus faecalis

US-09-634-341-29

Query Match

14.7%; Score 158.5; DB 20; Length 284;

Best Local Similarity 25.3%; Pred. No. 4.6e-07; Matches 56; Conservative 38; Mismatches 90; Indels 37; Gaps 7;

```
QY 1 MKKWTNRLMTIAGVLLIVA-----AYLFAKPHIDNYL-----HDKDKDE 40
  || | : : : : : || | : : : : :
Db 1 MMSKKRRRIIDGMILLIIGIGAFAYPFVSDALNNYLDQOIIAHYQAKASQENTKEMAE 60
  || | : : : : : || | : : : : :
QY 41 KIEQYDKNNKEQA-----SKDKQQAQAPQIPKDKSKVAGYIEIPDADIKPEVP 89
  || | : : : : : || | : : : : :
Db 61 LOEKMEKKNOELAKKGSNPGDLPFSETQKTKPKDKSYFESHIGVLTIPKINVRPIF- 119
  || | : : : : : || | : : : : :
QY 90 GPATPQOLNRGVSAFAENE---SLDDQNISIAHGTIFIDRPNYQFTMLKAAGKSMYFVK 146
  || | : : : : : || | : : : : :
Db 120 DKTNALLLEKGSLLGEGTSPGCTGTHAVISHRGLPQAKL-FTDLPKLKGDDEYIEY 178
  || | : : : : : || | : : : : :
QY 147 GNETRRYKMTSIRDVAPTDVGVLEQKGRDKQLTLITCDY 187
  || | : : : : : || | : : : : :
Db 179 NGKTLAYQVDQIKTVEPTDKDLHIESGD-LVTLLTCTPY 218
  || | : : : : : || | : : : : :
```

RESULT 11

US-60-223-804-29

; Sequence 29, Application US/60223804

; GENERAL INFORMATION:

; APPLICANT: Adderson, Elisabeth

; APPLICANT: Bohnsack, John

; TITLE OF INVENTION: GROUP B STREPTOCOCCUS POYPEPTIDES NUCLEIC ACIDS AND

; TITLE OF INVENTION: THERAPEUTIC COMPOSITIONS AND VACCINES THEREOF

; FILE REFERENCE: 2511-1-001P

; CURRENT APPLICATION NUMBER: US/60/223,804

; CURRENT FILING DATE: 2000-08-08

; NUMBER OF SEQ ID NOS: 37

; SOFTWARE: Patent In Ver. 2.0

; SEQ ID NO 29

; LENGTH: 284

; TYPE: PRF

; ORGANISM: Enterococcus faecalis

US-60-223-804-29

Query Match 14.7%; Score 158.5; DB 22; Length 284;

Best Local Similarity 25.3%; Pred. No. 4.6e-07;

Matches 56; Conservative 38; Mismatches 90; Indels 37; Gaps 7;

```
QY 1 MKKWTNRLMTIAGVLLIVA-----AYLFAKPHIDNYL-----HDKDKDE 40
  || | : : : : : || | : : : : :
Db 1 MMSKKRRRIIDGMILLIIGIGAFAYPFVSDALNNYLDQOIIAHYQAKASQENTKEMAE 60
  || | : : : : : || | : : : : :
QY 41 KIEQYDKNNKEQA-----SKDKQQAQAPQIPKDKSKVAGYIEIPDADIKPEVP 89
  || | : : : : : || | : : : : :
Db 61 LOEKMEKKNOELAKKGSNPGDLPFSETQKTKPKDKSYFESHIGVLTIPKINVRPIF- 119
  || | : : : : : || | : : : : :
QY 90 GPATPQOLNRGVSAFAENE---SLDDQNISIAHGTIFIDRPNYQFTMLKAAGKSMYFVK 146
  || | : : : : : || | : : : : :
Db 120 DKTNALLLEKGSLLGEGTSPGCTGTHAVISHRGLPQAKL-FTDLPKLKGDDEYIEY 178
  || | : : : : : || | : : : : :
QY 147 GNETRRYKMTSIRDVAPTDVGVLEQKGRDKQLTLITCDY 187
  || | : : : : : || | : : : : :
Db 179 NGKTLAYQVDQIKTVEPTDKDLHIESGD-LVTLLTCTPY 218
  || | : : : : : || | : : : : :
```

RESULT 12

US-09-292-437-6

; Sequence 6, Application US/09292437

; GENERAL INFORMATION:

; APPLICANT: Olaf Schneewind

; APPLICANT: Sarkis Mazmanian

; APPLICANT: Gwen Liu

; APPLICANT: Hung Ton-That

; TITLE OF INVENTION: IDENTIFICATION OF SORTASE GENE

; FILE REFERENCE: 510015.213

; CURRENT APPLICATION NUMBER: US/09/292,437

; CURRENT FILING DATE: 1999-04-15

; NUMBER OF SEQ ID NOS: 36

; SOFTWARE: FastSeq for Windows Version 3.0

; SEQ ID NO 6

; LENGTH: 284

; TYPE: PRF

; ORGANISM: Enterococcus faecalis

US-09-292-437-6

Query Match 14.7%; Score 158.5; DB 27; Length 284;

Best Local Similarity 25.3%; Pred. No. 4.6e-07;

Matches 56; Conservative 38; Mismatches 90; Indels 37; Gaps 7;

```
QY 1 MKKWTNRLMTIAGVLLIVA-----AYLFAKPHIDNYL-----HDKDKDE 40
  || | : : : : : || | : : : : :
Db 1 MMSKKRRRIIDGMILLIIGIGAFAYPFVSDALNNYLDQOIIAHYQAKASQENTKEMAE 60
  || | : : : : : || | : : : : :
QY 41 KIEQYDKNNKEQA-----SKDKQQAQAPQIPKDKSKVAGYIEIPDADIKPEVP 89
  || | : : : : : || | : : : : :
Db 61 LOEKMEKKNOELAKKGSNPGDLPFSETQKTKPKDKSYFESHIGVLTIPKINVRPIF- 119
  || | : : : : : || | : : : : :
QY 90 GPATPQOLNRGVSAFAENE---SLDDQNISIAHGTIFIDRPNYQFTMLKAAGKSMYFVK 146
  || | : : : : : || | : : : : :
Db 120 DKTNALLLEKGSLLGEGTSPGCTGTHAVISHRGLPQAKL-FTDLPKLKGDDEYIEY 178
  || | : : : : : || | : : : : :
QY 147 GNETRRYKMTSIRDVAPTDVGVLEQKGRDKQLTLITCDY 187
  || | : : : : : || | : : : : :
Db 179 NGKTLAYQVDQIKTVEPTDKDLHIESGD-LVTLLTCTPY 218
  || | : : : : : || | : : : : :
```

```

RESULT      13
US-09-134-000-6125
: Sequence 6125, Application US/09134000A
: GENERAL INFORMATION:
: APPLICANT: Lynn Doucette-Stamm et al
: TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ENTEROCOCCUS
: TITLE OF INVENTION: FAECALIS FOR DIAGNOSTICS AND THERAPEUTICS
: FILE REFERENCE: GFC-005
: CURRENT APPLICATION NUMBER: US/09/134,000A
: CURRENT FILING DATE: 1998-08-13
: NUMBER OF SEQ ID NOS: 6810
: SEQ ID NO 6125
: LENGTH: 315
: TYPE: PRF
: ORGANISM: Enterococcus faecalis
US-09-134-000-6125

```

Query Match	14.7%	Score 158.5	DB 15	Length 315
Best Local Similarity	25.3%	Pred. No. 5	3e-07	
Matches	56	Conservative	38	Mismatches 90
				Indels 37
				Gaps 7

```

OY      1 MKKMTNRLMTJAGVLLIVA-----ALFLAAPHIDNLT-----HDKRDE 40
Db      32 MKSKKKRRIIDGFMILLIITIGAFAPETYSDAINNTLDOOI IAHYAKKSQENTKMAE 91
OY      41 KIEQDKNVKEQA-----SKDRKQAKQEQIKRDKSKVAGYIEIPDADIKEEVVP 89
Db      92 LQEMKEKKNGCLAKGNSPGLDAPSEFQTKTKKRDKSYPESHITIGVLTIRKINRLTIF- 150
OY      90 GPATPEQLNGCVSAEENE---SLDDONISITAGHTTIDRENYOFTNLKAAKSGMYEYK 146
Db      151 DKTALLLEKSSSLLEGTSYPTGGSTNHAVISGHRGIPQAKL-FTDPELPLKGGDEFEYEV 209
OY      147 GNETRRYKYSITRDVKPDPVDELDQKQKQKLLITICDXY 187
Db      210 NGKTLAYOVDDIKIVPEPTDKDLHIESGQD-LVTLTTCCTP 249

```

```

1 RESULT 14
2 US-09-107-532-5472
3 ; Sequence 5472, Application US/09107532
4 ;
5 ; GENERAL INFORMATION:
6 ; APPLICANT: Lynn A Doucellette-Stamm and David Bush
7 ; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO
8 ; TITLE OF INVENTION: ENTEROCOCCUS FACIUM FOR
9 ; TITLE OF INVENTION: DIAGNOSTICS AND THERAPEUTICS
10 ; NUMBER OF SEQUENCES: 7308
11 ;
12 ; CORRESPONDENCE ADDRESS:
13 ; ADDRESSEE: GENOME THERAPEUTICS CORPORATION
14 ; STREET: 100 Beaver Street
15 ; City: Waltham
16 ; STATE: Massachusetts
17 ; COUNTRY: USA
18 ;
19 ; ZIP: 02354
20 ;
21 ; COMPUTER READABLE FORM:
22 ; MEDIUM TYPE: CD/ROM ISO9660
23 ;
24 ; COMPUTER:
25 ; OPERATING SYSTEM:
26 ;
27 ; SOFTWARE:
28 ; CURRENT APPLICATION DATA:
29 ; APPLICATION NUMBER: US/09/107,532
30 ; FILING DATE:
31 ;
32 ; PRIOR APPLICATION DATA:
33 ; APPLICATION NUMBER: 60/ 085598
34 ; FILING DATE: May 14, 1998
35 ;
36 ; PRIOR APPLICATION DATA:
37 ; APPLICATION NUMBER: 60/051571
38 ; FILING DATE: July 2, 1997
39 ;
40 ; ATTORNEY/AGENT INFORMATION:
41 ; NAME: Atinidello, Pamela Deneke

```

```

1 REGISTRATION NUMBER: 40_489
2 REFERENCE/DOCKET NUMBER: GTC-012
3 TELECOMMUNICATION INFORMATION:
4 TELEPHONE: (781)893-5007
5 TELEFAX: (781)893-8277
6 INFORMATION FOR SEQ ID NO: 5472:
7 SEQUENCE CHARACTERISTICS:
8     LENGTH: 259 amino acids
9     TYPE: amino acid
10    TOPOLOGY: linear
11    MOLECULE TYPE: protein
12    HYPOTHEetical: YES
13    ORIGINAL SOURCE:
14    ORGANISM: Enterococcus faecium
15    FEATURE:
16    NAME/KEY: misc_feature
17    LOCATION: 1...259
18 OS-09-107-532-5472

```

Query Match	14.6%	Score 157.5	DB 15	length 259
Best Local Similarity	25.4%	Pred. No. 5e-07		
Matches 57, Conservative	39	Mismatches	89	Indels 39
				Gaps 9

[illegible]

```

RESULT 15
US-09-292-437-4
: Sequence 4, Application US/09292437
: GENERAL INFORMATION:
: APPLICANT: Olaf Schneewind
: APPLICANT: Sarkis Mazmanian
: APPLICANT: Gwen Liu
: APPLICANT: Hung Ton-That
: TITLE OF INVENTION: IDENTIFICATION OF SORTASE GENE
: FILE REFERENCE: 510015.213
: CURRENT APPLICATION NUMBER: US/09/292,437
: CURRENT FILING DATE: 1999-04-15
: NUMBER OF SEQ ID NOS: 36
: SOFTWARE: FASTSEQ for Windows Version 3.0
: SEQ ID NO 4
: LENGTH: 227
: TYPE: PRF
: ORGANISM: Streptococcus pyogenes
US-09-292-437-4

```

Query March	14.5%	Score 156;	DB 27;	Length 227;
Best Local Similarity	32.6%	Pred. No. 5.8e-07;		
Matches 47;	Conservative 19;	Mismatches 60;	Indels 18;	Gaps 7;

QY 56 DKKOARPOITKOKSKVAGYEIEIPDAIDIKEPYGPATPQLNRGVSFAENESLDDONT 115
|||
|||
Db 32 DKNTSELTÖI--ENNIDMGIVEPSIKVTPLTY-HYTTEDELTIRKGAG----HLESGAL 82
|||
QY 116 SIAC--HTFIDR---PNYO-FTNLKAARKGSKVYERKVENERYKKMYTISRDVKPTDVG 167
|||
|||
Gb 83 PVGGDGTHTVISAHRGLPSASMEFTNMLNVKGDGFYFRVLNKLYAVVDOLILVEPOGYT 142
|||

OY 168 VIDEQKDKQLTLITCDDYNEXT 191
| | | | | | | | | |
Db 143 SLGVMGKD-YATLVCTPYGVNT 165

Search completed: March 2, 2001, 10:24:18
job time: 125 sec

THIS PAGE BLANK (USPTO)